

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 3744.31 Seconds
(without alignments)
9989.844 Million cell updates/sec

Title: US-10-082-830-98
Perfect score: 863
Sequence: 1 atgggttggtcattcaact.....gtgagaccaatgatagctg 863

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_on.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	613	71.0	162100	9	AC093297 Homo sapi
C 2	613	71.0	176269	2	AC012663 Homo sapi
C 3	72.2	8.4	154837	2	AL845425 Danio rer
C 4	72.2	8.4	161204	2	AL845324 Danio rer
C 5	71	8.2	183662	2	AC113205 Mus muscu
C 6	69.8	8.1	151406	5	AL935208 Zebrafish
C 7	67.6	7.8	6062	5	AL251789 Sequence
C 8	67.6	7.8	66993	2	AC138074 Homo sapi
C 9	67	7.8	146871	2	EX842689 Danio rer
C 10	66.2	7.7	6195	6	AX345493 Sequence
C 11	65.6	7.6	239339	2	EX548071 Danio rer
C 12	65.4	7.6	1407	8	AJ592026 Arabidops
C 13	65.4	7.6	1434	8	AJ592058 Arabidops
C 14	65.4	7.6	3683	6	AX598999 Sequence
C 15	65.2	7.6	1453	8	AJ591978 Arabidops
C 16	64	7.4	54345	3	AC084152 Caenorhab
C 17	63.6	7.4	151900	9	AC107419 Homo sapi
C 18	63.6	7.4	349980	6	AX344553 Sequence
C 19	63.6	7.4	349980	6	AX344554 Sequence
C 20	63.4	7.3	61864	3	CEY5088A Arabidops
C 21	63	7.3	1228	8	AJ592059 Arabidops
C 22	63	7.3	172816	9	AC093899 Homo sapi
C 23	62.8	7.3	7128	6	AX346461 Sequence
C 24	62.8	7.3	40324	6	AX458634 Sequence
C 25	62.6	7.3	104992	2	AC005504 Plasmodiu
C 26	62.6	7.3	169546	2	AC004157 Plasmodiu
C 27	62.6	7.3	250421	3	AE014849 Plasmodiu
C 28	62.4	7.2	93791	2	AC138073 Homo sapi
C 29	62.4	7.2	161699	9	AC107300 Homo sapi
C 30	62.2	7.2	11147	6	AX598954 Sequence
C 31	62.2	7.2	14147	6	AX767512 Sequence
C 32	62.2	7.2	14147	6	AX795870 Sequence
C 33	62.2	7.2	14147	6	AX822390 Sequence
C 34	62.2	7.2	14147	6	AX826030 Sequence
C 35	62.2	7.2	136335	8	AC005275 Arabidops
C 36	62.2	7.2	195429	8	ATCHRIV8 Arabidops
C 37	61.8	7.2	77242	5	AL591462 Zebrafish
C 38	61.8	7.2	131682	9	AL672277 Human DNA
C 39	61.8	7.2	13506	9	AC009950 Homo sapi
C 40	61.6	7.1	12029	3	AE001372 Plasmodiu
C 41	61.4	7.1	1184	8	AJ591982 Arabidops
C 42	61.4	7.1	157661	2	EX470134 Danio rer
C 43	61.4	7.1	192182	5	EX649602 Zebrafish
C 44	61.4	7.1	254050	3	PFA929358 Plasmodiu
C 45	60.8	7.0	174273	9	AL162759 Human chr

ALIGNMENTS

RESULT 1
AC093297/c AC093297 162100 bp DNA linear PRI 27-MAR-2002
LOCUS Homo sapiens chromosome 5 clone RP11-53019, complete sequence.
DEFINITION AC093297
ACCESSION AC093297.3 GI:19747150
VERSION AC093297.3
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162100)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished

Pred. No. is the number of results predicted by chance to have a

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168178 bases at least Q40
Consensus quality: 172094 bases at least Q30
Consensus quality: 173552 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 174969; sum-of-contigs
Quality coverage: 6.0 in Q20 bases; agarose-fp
Quality coverage: 5.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1315: contig of 1315 bp in length
1316 1415: gap of 100 bp
1416 2587: contig of 1172 bp in length
2588 2687: gap of 100 bp
2688 2947: contig of 260 bp in length
2948 3047: gap of 100 bp
3048 4582: contig of 1535 bp in length
4583 4682: gap of 100 bp
4683 8407: contig of 3725 bp in length
8408 16350: contig of 7843 bp in length
16351 16450: gap of 100 bp
16451 21745: contig of 5295 bp in length
21746 32523: contig of 10684 bp in length
32524 32623: gap of 100 bp
32624 50188: contig of 17559 bp in length
50189 50288: gap of 100 bp
50289 72435: contig of 22147 bp in length
72436 72535: gap of 100 bp
72536 93128: contig of 20593 bp in length
93129 93228: gap of 100 bp
93229 114608: contig of 21380 bp in length
114609 114708: gap of 100 bp
114709 142449: contig of 27741 bp in length
142450 142549: gap of 100 bp
142550 176269: contig of 33720 bp in length.
```

FEATURES
Location/Qualifiers
1..176269
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-3B17"
/clone_lib="RPC1-11 Human Male BAC"
1..1315
/note="assembly_fragment"
1416..2587
/note="assembly_fragment"
2688..2947
/note="assembly_fragment"
clone_end:T7
vector_side:right
3048..4582
/note="assembly_fragment"
4583..8407
/note="assembly_fragment"
8508..16350
/note="assembly_fragment"
16451..21745
/note="assembly_fragment"
clone_end:SP6
vector_side:right
21846..32529
/note="assembly_fragment"

```
misc_feature 32630..50188  
/note="assembly_fragment"  
misc_feature 50289..72435  
/note="assembly_fragment"  
misc_feature 72536..93128  
/note="assembly_fragment"  
misc_feature 93229..114608  
/note="assembly_fragment"  
misc_feature 114709..142449  
/note="assembly_fragment"  
misc_feature 142550..176269  
/note="assembly_fragment"  
  
ORIGIN  
Query Match 71.0%; Score 613; DB 2; Length 176269;  
Best Local Similarity 72.8%; Pred. No. 9e-97;  
Matches 625; Conservative 0; Mismatches 231; Indels 2; Gaps 1;  
  
QY 1 ATGGGTTGGTCATTCACACTTTAAGTTTATGAGATATATCCATGTTGATTTGTAGCTG 60  
Db |||||  
QY 99193 ATGGGTTGGTCATTCACACTTTAAGTTTATGAGATATATCCATGTTGATTTGTAGCTG 99134  
Db |||||  
QY 61 TGGTTTGATGATTTTACATTATATAGTATTACATTCCATGATAGTTTCTCAGTAGATAA 120  
Db |||||  
QY 99133 TGGTTTGATGATTTTACATTATATAGTATTACATTCCATGATAGTTTCTCAGTAGATAA 99074  
Db |||||  
QY 121 TCTCTCTATTGTTAAACATTTTGGCTGCTCTCTCATTTTGACCTATTATAACAGCCCTCT 180  
Db |||||  
QY 99073 TCTCTCTATTGTTAAACATTTTGGCTGCTCTCTCATTTTGACCTATTATAACAGCCCTCT 99014  
Db |||||  
QY 181 TTAATAATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGAATAAAAGTGGCT 240  
Db |||||  
QY 99013 TTAATAATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGAATAAAAGTGGCT 98954  
Db |||||  
QY 241 GAATCATCTTCAACTCTAGTAGTAGTCAAACTGTTTCTAAAAGTGTTTTATTTAACT 300  
Db |||||  
QY 98953 GAATCATCTTCAACTCTAGTAGTAGTCAAACTGTTTCTAAAAGTGTTTTATTTAACT 98894  
Db |||||  
QY 301 ATTATATGTCATTTTGAACAGCTCTTCACTTACTAGCAATTTTATATCAGCAACACTT 360  
Db |||||  
QY 98893 ATTATATGTCATTTTGAACAGCTCTTCACTTACTAGCAATTTTATATCAGCAACACTT 98834  
Db |||||  
QY 361 GTTATTGTGACACTTTTAAAGTTTCACTCACTGGNNNNNNNNNNNNNNNNNNNNNNNN 420  
Db |||||  
QY 98833 GTTATTGTGACACTTTTAAAGTTTCACTCACTGG--GTAACATGACACTTTTGTGCTT 98776  
Db |||||  
QY 421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480  
Db |||||  
QY 98775 CTAAATTGCATTTCATGCTACTAAATGAAATTTGAGCTTGTGTTTAGTACAATTTATGCTC 98716  
Db |||||  
QY 481 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540  
Db |||||  
QY 98715 ATTTTATTCTCTGCTGTGAAATGCCCTGTTGAGGATTTGTTCTCATCTTTTAAATA 98656  
Db |||||  
QY 541 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600  
Db |||||  
QY 98655 GTTTCAAAATATGTTGCCCTAGCTTGTGGCTTGACCTTTTTCAGTTTCTTCAGTATTACATTT 98596  
Db |||||  
QY 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660  
Db |||||  
QY 98595 GATTAAGAGAGAGTCTTAAACTTCAATGTGATCATTTTATGTTTGTCTCATTTTAAATA 98536  
Db |||||  
QY 661 ACCTGGTTTAAATATTTTCAAAATCGACATTATGATATATATCTTCCAAAAATTTTAAATA 720  
Db |||||  
QY 98535 ACCTGGTTTAAATATTTTCAAAATCGACATTATGATATATATCTTCCAAAAATTTTAAATA 98476  
Db |||||  
QY 721 TTTTGTCTTTTTCACATTTTGTAGCTTTAGCTGAGTGGATTCATTTCTGTGTGTGTG 780  
Db |||||  
QY 98475 TTTTGTCTTTTTCACATTTTGTAGCTTTAGCTGAGTGGATTCATTTCTGTGTGTGTG 98416  
Db |||||  
QY 781 TGAGATAAGTCTTTTTCATGTTTTCCTATGAATAAAATTTATTCCTTCTGTATTGCGAG 840  
Db |||||  
QY 98415 TGAGATAAGTCTTTTTCATGTTTTCCTATGAATAAAATTTATTCCTTCTGTATTGCGAG 98356  
Db |||||
```

```

Oy      841 GTAGCTGAGACCAATGAT 858
Db      98355 GTAGCTGAGACCAATGAT 98338

RESULT 3
AL845425/c
LOCUS   AL845425      154837 bp    DNA    linear    HTG 15-NOV-2002
DEFINITION Danio rerio clone CH211-123B21, WORKING DRAFT SEQUENCE, 19
unordered pieces.
ACCESSION AL845425.3 GI:25137035
VERSION   AL845425.3
KEYWORDS  HTG; HTGS, PHASE1; HTGS DRAFT.
SOURCE   Danio rerio (zebrafish)
ORGANISM Danio rerio
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
          Burton, J.
REFERENCE 1 (bases 1 to 154837)
AUTHORS  Direct Submission
TITLE    Submitted (17-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL  Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
COMMENT  Clone requests: clonerequest@sanger.ac.uk
          On Nov 19, 2002 this sequence version replaced gi:22416248.
          ----- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zface@sanger.ac.uk
          ----- Project Information
          Center project name: zcl13B21
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 149574 bases at least Q40
          Consensus quality: 151215 bases at least Q30
          Consensus quality: 152331 bases at least Q20
          Insert size: 153037; sum-of-contigs
          Insert size: 168327; 8.3% error; agarose-fp
          Quality coverage: 4.53x in Q20 bases; sum-of-contigs Quality
          coverage: 4.24x in Q20 bases; agarose-fp
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 19 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 1 7902: contig of 7902 bp in length
          * 7903 8002: gap of 100 bp
          * 8003 18799: contig of 10797 bp in length
          * 18800 18899: gap of 100 bp
          * 18900 32870: contig of 13971 bp in length
          * 32871 32970: gap of 100 bp
          * 32971 37989: contig of 5019 bp in length
          * 37990 38089: gap of 100 bp
          * 38090 40307: contig of 2218 bp in length
          * 40308 40407: gap of 100 bp
          * 40408 51942: contig of 11535 bp in length
          * 51943 52042: gap of 100 bp
          * 52043 56846: contig of 4804 bp in length
          * 56847 56947: gap of 100 bp
          * 56948 63028: contig of 6079 bp in length
          * 63029 63126: gap of 100 bp
          * 63127 77084: contig of 13959 bp in length
          * 77085 77185: gap of 100 bp
          * 77186 79709: contig of 2525 bp in length
          * 79710 79809: gap of 100 bp
          * 79810 84969: contig of 5160 bp in length

          85069: gap of 100 bp
          89945: contig of 4876 bp in length
          90045: gap of 100 bp
          104918: contig of 14873 bp in length
          105018: gap of 100 bp
          103886: contig of 4868 bp in length
          109887: gap of 100 bp
          120755: contig of 10769 bp in length
          120855: gap of 100 bp
          133958: contig of 13103 bp in length
          134058: gap of 100 bp
          141783: contig of 7725 bp in length
          141883: gap of 100 bp
          147865: contig of 5982 bp in length
          147965: gap of 100 bp
          154837: contig of 6872 bp in length.

FEATURES
            Location/Qualifiers
                1..154837
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7955"
                    /clone_lib="CH211-123B21"
                    /clone_lib="CHORI-211"
                1..7902
                    /note="assembly fragment:00270"
                    fragment_chain:1
                8003..18799
                    /note="assembly fragment:00742"
                    fragment_chain:1
                18900..32870
                    /note="assembly fragment:00868"
                    fragment_chain:1
                32971..37989
                    /note="assembly fragment:00250"
                    fragment_chain:1
                38090..40307
                    /note="assembly fragment:00721"
                    fragment_chain:1
                40408..51942
                    /note="assembly fragment:01237"
                    fragment_chain:2
                52043..56846
                    /note="assembly fragment:00594"
                    fragment_chain:2
                56947..63025
                    /note="assembly fragment:00160"
                    fragment_chain:2
                63126..77084
                    /note="assembly fragment:00251"
                    fragment_chain:2
                77185..79709
                    /note="assembly fragment:01322"
                    fragment_chain:3
                79810..84969
                    /note="assembly fragment:00831"
                    fragment_chain:3
                85070..89945
                    /note="assembly fragment:00266"
                    fragment_chain:3
                90046..104918
                    /note="assembly fragment:00079"
                    fragment_chain:3
                105019..109886
                    /note="assembly fragment:00272"
                    fragment_chain:4
                109987..120755
                    /note="assembly fragment:00186"
                    fragment_chain:4
                120856..133958
                    /note="assembly fragment:00513"
                    fragment_chain:4
                134059..141793
                    /note="assembly fragment:01199"
                    fragment_chain:4

```



```
misc_feature 141884..147865
/notes="assembly fragment:00008
fragment_chain:5"
misc_feature 147966..154837
/notes="assembly fragment:00289
fragment_chain:5
clone_end:SP6
vector_side:right"
ORIGIN
Query Match 8.4%; Score 72.2; DB 2; Length 154837;
Best Local Similarity 32.3%; Pred. No. 0.0013;
Matches 260; Conservative 0; Mismatches 546; Indels 0; Gaps 0;
Qy 20 TTTAAGTTTATGAGATATATCCAGTGTGAATTTTGTAGCTGGCTTGATGATTTTACA 79
Db 145649 TTTTATTTTATTTCTATTTTCAATGTTTCTCTTTTAAATTAATATATATATATATAA 145590
Qy 80 TTATATAGTATTACATTCATCGATAGTCTCTCAGTAGATATCTCTCTATGTTTAAACAT 139
Db 145589 TTATATATATATATATATATTAACCTTTTATTTATTTATTTATTTATTTGCTTTACT 145530
Qy 140 TTGGTTTGTCTTCATTTTGCACCTATTTTAAACAGGCTCTTTAAATATACATCTTCTCG 199
Db 145529 TTTTGTATTTATTTGATCTTTTACTTTTACTTTTATTTATTTATTTTATTTATCTGTTT 145470
Qy 200 TAGTGATGCTAGAAATGGAGTGGCTGGAATAAAGTGGCTGATCATCTTCAACTCTAG 259
Db 145469 TTATTTGTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 145410
Qy 260 TAAGATGCTAAACCTGTTTCTTAAAGTGTTTATTTTAACTATTTATATGTCATTTTAA 319
Db 145409 TACTCTTTTACTTTTATTTTATTTTATTTTATTTTATTTTCTTCTTCTTTTATTTAT 145350
Qy 320 CAGCTCTTTCACCTACTAGCAATTTATTTACGACACACTTGTATGTCAGACTTTTAA 379
Db 145349 TATTTAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 145290
Qy 380 GTTTTCTTCACTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 439
Db 145289 TAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 145230
Qy 440 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 499
Db 145229 CTCTTACTTTTACTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTACT 145170
Qy 500 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 559
Db 145169 TTTTACTTTTATTTAATATCAATTTATTTTATTTTATTTTATTTTATTTTATTTTCTT 145110
Qy 560 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 619
Db 145109 TTTCTTTTAGATTTTTCAGATATTTTCTTCTATTTTCTATTTTATTTTATTTTATTT 145050
Qy 620 NNNNNNNNGTGCATATTTATGTTTGTCTCAITTAATAAAACCTGTTAAATTTTCA 679
Db 145049 TATTTCAATTTATTTTATTTTATGTTTACTTTTATTTTATTTTATTTTATTTAT 144990
Qy 680 AAATCGACATTTATCATATTTATCTTCCAAAATTTAATAATTTGCTTTTTCACAT 739
Db 144989 ATTTATTTTACTTAAATATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 144930
Qy 740 TTAGCTTTTACTGCTGGAATTTCTATTTCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 799
Db 144929 ATTTTATTTTACTTTTACTTTTACTTTTATTTTATTTTATTTTATTTTATTTAT 144870
Qy 800 GTTTTTCCTCATGAATAAATTTT 825
Db 144869 TTTATTTTATTTTATTTTATTTTATTT 144844
RESULT 4
```

```
AL845324
LOCUS
DEFINITION
AL845324
ACCESSION
AL845324
VERSION
AL845324.15
GI:35209193
KEYWORDS
HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 161204)
Direct Submission
Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfisch-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 24, 2003 this sequence version replaced gi:31616709.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfisch-help@sanger.ac.uk
----- Project Information
Center project name: zc122114
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 161092 bases at least Q40
Consensus quality: 161096 bases at least Q30
Consensus quality: 161100 bases at least Q20
Insert size: 161104; sum-of-contigs
Insert size: 166733; 4.8% error; agarose-fp
Quality coverage: 14.82x in Q20 bases; sum-of-contigs Quality
coverage: 14.32x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 140163: contig of 140163 bp in length
* 140164 140263: gap of 100 bp
* 140264 161204: contig of 20941 bp in length.
FEATURES
Location/Qualifiers
1..161204
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-122L14"
/clone_lib="CHORI-211"
misc_feature 1..140163
note="assembly fragment:01479
fragment_chain:1"
misc_feature 140264..161204
note="assembly fragment:04647
fragment_chain:1"
ORIGIN
Query Match 8.4%; Score 72.2; DB 2; Length 161204;
Best Local Similarity 32.3%; Pred. No. 0.0013;
Matches 260; Conservative 0; Mismatches 546; Indels 0; Gaps 0;
Qy 20 TTTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGGCTTGATGATTTTACA 79
Db 9118 TTTTATTTTATTTCTATTTTCAATGTTTCTCTTTTAAATTAATATATATATATAA 9177
Qy 80 TTATATAGTATTACATTCATCGATAGTCTCTCAGTAGATATCTCTCTATGTTTAAACAT 139
```



```

* provided by the submittor
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 1269: contig of 1269 bp in length
* 1270 1369: gap of 100 bp
* 1370 3799: contig of 2430 bp in length
* 3800 3899: gap of 100 bp
* 3900 16887: contig of 12988 bp in length
* 16888 16987: gap of 100 bp
* 16988 44512: contig of 27525 bp in length
* 44513 44612: gap of 100 bp
* 44613 110938: contig of 66226 bp in length
* 110939 156050: contig of 45112 bp in length
* 156051 156350: gap of 100 bp
* 156151 183662: contig of 27512 bp in length.
*
FEATURES
      source
      1..183662
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /clone_lib="RP23-319G18"
      /clone_lib="RPC1-23 Female Mouse BAC"
      1..1269
      /note="assembly_fragment
      clone_end:SP6
      vector_side:left"
      1370..3799
      /note="assembly_fragment"
      3900..16887
      /note="assembly_fragment"
      16988..44512
      /note="assembly_fragment"
      44613..110938
      /note="assembly_fragment"
      110939..156050
      /note="assembly_fragment"
      156151..183662
      /note="assembly_fragment
      clone_end:T7
      vector_side:right"

ORIGIN

Query Match      8.2%; Score 71; DB 2; Length 183662;
Best Local Similarity 35.7%; Pred. No. 0.002;
Matches 196; Conservative 0; Mismatches 334; Indels 19; Gaps 4;

QY      313 TTTTGACAGCTCTTTCACCTACTAGCAATATTATATCAGCAACTGTTATTGTGAGA 372
         |||||
Db       57070 TTTCAAATAGTTTGTGTTAACTAGCAATAGTCAGTCACACACTTGTACTCTAAGC 57011

QY      373 CTTTAAAGTTTTCATTCACTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 432
         |||||
Db       57010 CTTTAAAGGTTTTCCTCATTTGCTATAAAATAGCTATTCTCTCTGCTGTTCTATGGT 56951

QY      433 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 492
Db       56950 ATAATTCCCTTCTTGCTAAATGAAATTTAGAAAAGTTTCTATTCAATAGTCATTTCTCTTC 56891

QY      493 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 552
Db       56890 TGTGTGGAATTCAGTTTAAGATTTTGTTCATCTTTGCACGTGATTTTCAACACAGATTT 56831

QY      553 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 612
Db       56830 ACTAGTTTGTGAATTTACCTTTTCCACCTACTCTATTATTATCAATTCATTTGGGAGAACATCT 56771

QY      613 NNNNNNNNNNNNNNGTGATCATATTATTTAGTTTGTCTATTATAAAAAACCTGGTTAAAT 672
Db       56770 TCAAGT-----TGGTCATCATATGCTGATCTCATTTTAAAAAATTACGATAGG 56721

QY      673 ATTTAC--AAATCGACATTATGATATATATCTTCCAAAAATTTTAATAATTTTGTCTTT 730

```

```

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-241N17"
/clone_lib="CHORI-211"

ORIGIN
Query Match      8.1%; Score 69.8; DB 5; Length 151406;
Best Local Similarity 32.6%; Pred. No. 0.0033;
Matches 250; Conservative 0; Mismatches 515; Indels 1; Gaps 1;

QY 1 ATGGGTTGGTCATCACTAACTTAAGTTTATGAGATATATCCATGGATAGTCTCAGTAGATAA 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 TGGTTTGATGATTTTACATATATAGTATATACATTCATCCATGGATAGTCTCAGTAGATAA 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68035 TTATATTTTATTAATCAATTTATTAATATATATTTATTTATTTATTTTGTGTTT 68094
QY 121 TCCCTCATGTTTAAAGATTCGCTTCCTCTCATTTTGAACCTATTTTAAACAGGCTCT 180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68095 TATTCGATTTCTTTAAATTAATTTAAATATATTTTATTTTATTTTATTTTAAATAT 68154
QY 181 TTAATATACACTTCTGAGTGTATGCTAGAAATGAGTGGCTGGAATAAAGTGCT 240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68155 TTTATTTCTTAAATTCATATTTATTTAAATTAATTAATTAATTTTGTGTTAATTTAT 68214
QY 241 GAATCATCTTCACTAGTAAAGTCAAACTGTTTTCTAAA- GTGTTTTATTTTAAAC 299
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68215 TTATATCTTTTATTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTAT 68274
QY 300 TATTATATGCAATTTGAACAGCTCTTCACTACTAGCAATTTATATCAGCAACT 359
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68275 TTTTAAATCTCTTTGTTTAACTATTTATTTTAAATTTATTTTATTTTAAATATTTT 68334
QY 360 TGTATTTGTCAGACTTTAAAGTTTTCATTCAGTGGNNNNNNNNNNNNNNNNNNNNNN 419
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68335 TATTTTGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 68394
QY 420 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 479
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68395 AATATTTATTTTGTATTTATTTTAAATTTCTTTTATTTATTTATTTATTTATTT 68454
QY 480 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 539
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68455 TTTTCTCTTAAATTTGTTTATTTAAATTTATTTATTTATTTATTTATTTATTTAT 68514
QY 540 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 599
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68515 TTTATTTTCATAATTTTCA-TATTTTATTTAAATTTTATTTATTTATTTATTTTGA 68574
QY 600 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 659
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68575 TTTATTTATTTCAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAT 68634
QY 660 AACCTGTTAAATTTTCAAAATCGACATTTATGATATATTTCTCCAAAATTTTAAATA 719
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68635 ATTTTAGGATTTTTCATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTA 68694
QY 720 ATTTTGTCTTTTTCATTTTATGCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 765
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68695 ATTTTATATAATTTTATTTACTTTTATTTACTTTTATTTACTTTTATTTACTTTTAT 68740

RESULT 7
AX251789 LOCUS AX251789 5062 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 50 from Patent WO0168911.
ACCESSION AX251789
VERSION AX251789.1 GI:15985144
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM synthetic construct
```

```

artificial sequences.
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the cell cycle
JOURNAL Patent: WO 0168911-A 50 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source 1. 6062
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match      7.8%; Score 67.6; DB 6; Length 6062;
Best Local Similarity 32.8%; Pred. No. 0.018;
Matches 270; Conservative 0; Mismatches 552; Indels 1; Gaps 1;

QY 20 TTTAAGTTTATGAGATATATCCATGTTGGAATTTTGTAGCTGCTGTTGATGATTTTACA 79
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5122 TTTAGTCGTCGTATATATATAGTAATTTTAAATGAGTTTGTAGTTAGTTTATTTT 5181
QY 80 TTATATAGTATTACATTCATGGATAGTCTCAGTAGATAATCTCTATTTGTTTAAACAT 139
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5182 TTGATTATAAATTTATATGTTTATGTTTAAATATTTGTTTGGATGTTTAAAGAT 5241
QY 140 TTGGCTTCCTCTCATTTTGTACCTATTTTAAACAGGCTCTTTAAATATACACTTCTCTG 199
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5242 ATTTTAAAAATTTTAAATTTTATTTTATTTATTTATTTATTTTATTTTAAATTTTAA 5301
QY 200 TAGCTATGCTAGAAATGAGTGGCTGGAATAAAGTGGCTGAATCATCTCAACTCTAG 259
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5302 AAGTGGGTATTATTAAT-TCGTTTGTGAAATAGAAATAGGTAATTTATTTAGTTCGT 5360
QY 260 TAAGATGTCAAACTGTTTCTTAAAGTGTTTTATTTTAACTATTATATATGCAATTTTCAA 319
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5361 TTTTGTGTTAGTTATTTAGAAAGTTTATTTAGTTTATTTTATTTTAAATTTTAA 5420
QY 320 CAGCTCTTCACTACTAGCAATTTATATCAGCAACTGTTTATTTGTCAGACTTTTAA 379
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5421 TTTTGTGTTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 5480
QY 380 GTTTTCATTCAGTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 439
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5481 GTAGTAGTTTGTGTCGGTAAATTTGTAATTTATTTTATATCTGATTTAGAGTGAT 5540
QY 440 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 499
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5541 TTTTAAATTTTAAATTTATTTAGTTTGTGTTTAAAGTTTAAATTTTAAAGTAA 5600
QY 500 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 559
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5601 ATTTGTATGATATATTTTATTTTGTGTTAAATTTGTTATATTTGTTTATTTTAAAT 5660
QY 560 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 619
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5661 TTAGTAGTTTGTGTTTATTTATTTTATTTTATTTATTTATTTATTTATTTATTTG 5720
QY 620 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 679
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5721 TTTTAAAGTTTGTAGTTTAAATTTATTTTAAAGTTTAAAGTTTAAAGTTTAAAGTT 5780
QY 680 AAATCGACATTTATGATATTTATCTCCAAAATTTTAAATTTTGTCTTTTTCACATT 739
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5781 AATTAGGTAGGTATTTTATTTTATTTTATAGTATTTTGTGTTTATTTATGTTAGT 5840
QY 740 TTAGTCTTTAGCTCAGTGGAAATTCATTTCTGTGTTGTTGTTGTTGTTGTTGTTT 799
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5841 TAGTAGTAGAGTTTGTGAAATTTTGTGTTTATTTTAAAGAGATAAATTTAAGTTTAA 5900
QY 800 GTTTTCCCTATGAATAAATTTTCTCTCTGTTATTTGTCAGGT 842
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5901 TTTTAAAGTTAAGATTTTGTGTTTATTTGTTTATTTTGTAGTT 5943
```

```

RESULT 8
AC138074
LOCUS
DEFINITION
AC138074 Homo sapiens chromosome 18 clone RP13-907D8 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION
AC138074.1 GI:26553381
VERSION
HTG; HTGS PHASE0.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 66993)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP13-907D8
Unpublished
2 (bases 1 to 66993)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Chepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Landblad-Ton,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Ronan,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
TITLE
Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
All repeats were identified using RepeatMasker:
COMMENT
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L28874
Center clone name: 907_D_8
-----
* NOTE: This record contains 55 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 1116: contig of 1116 bp in length
* 1117 1216: gap of 100 bp
* 1217 2336: contig of 1120 bp in length
* 2337 2436: gap of 100 bp
* 2437 3561: contig of 1125 bp in length
* 3562 3661: gap of 100 bp
* 3662 4781: contig of 1120 bp in length
* 4782 4881: gap of 100 bp
* 4882 6001: contig of 1120 bp in length
* 6002 6101: gap of 100 bp
*
* 7244: contig of 1143 bp in length
* 7344: gap of 100 bp
* 8458: contig of 1114 bp in length
* 8459 gap of 100 bp
* 8559 9640: contig of 1082 bp in length
* 9641 9740: gap of 100 bp
* 10846: contig of 1106 bp in length
* 10847 gap of 100 bp
* 10947 12103: contig of 1157 bp in length
* 12104 12203: gap of 100 bp
* 12204 13200: contig of 1117 bp in length
* 13321 13420: gap of 100 bp
* 13421 14520: contig of 1100 bp in length
* 14521 14620: gap of 100 bp
* 14621 15773: contig of 1153 bp in length
* 15774 15873: gap of 100 bp
* 15874 16991: contig of 1118 bp in length
* 16992 17091: gap of 100 bp
* 17092 18240: contig of 1149 bp in length
* 18241 18340: gap of 100 bp
* 18341 19476: contig of 1136 bp in length
* 19477 19576: gap of 100 bp
* 20721 20821: contig of 1145 bp in length
* 20822 21942: contig of 1121 bp in length
* 21943 22042: gap of 100 bp
* 22043 23141: contig of 1099 bp in length
* 23142 23241: gap of 100 bp
* 23242 24372: contig of 1131 bp in length
* 24373 25627: contig of 1155 bp in length
* 25628 25727: gap of 100 bp
* 25728 26888: contig of 1161 bp in length
* 26889 gap of 100 bp
* 28030: contig of 1042 bp in length
* 28031 28131: gap of 100 bp
* 28132 29303: contig of 1173 bp in length
* 29304 29403: gap of 100 bp
* 30554 30554: contig of 1151 bp in length
* 30555 30654: gap of 100 bp
* 31751: contig of 1097 bp in length
* 31752 31851: gap of 100 bp
* 31852 32966: contig of 1115 bp in length
* 32967 33066: gap of 100 bp
* 33067 34113: contig of 1047 bp in length
* 34114 34213: gap of 100 bp
* 35329: contig of 1116 bp in length
* 35330 35429: gap of 100 bp
* 35430 36574: contig of 1145 bp in length
* 36575 36674: gap of 100 bp
* 36675 37828: contig of 1154 bp in length
* 37829 37928: gap of 100 bp
* 37929 39063: contig of 1135 bp in length
* 39064 39163: gap of 100 bp
* 40289: contig of 1126 bp in length
* 40290 40389: gap of 100 bp
* 40390 41471: contig of 1082 bp in length
* 41472 41571: gap of 100 bp
* 41572 42634: contig of 1063 bp in length
* 42635 42734: gap of 100 bp
* 42735 43904: contig of 1170 bp in length
* 43905 44004: gap of 100 bp
* 44005 45128: contig of 1124 bp in length
* 45129 45228: gap of 100 bp
* 45229 46358: contig of 1130 bp in length
* 46359 46458: gap of 100 bp
* 46459 47595: contig of 1137 bp in length
* 47596 47695: gap of 100 bp
* 47696 48785: contig of 1090 bp in length
* 48786 48885: gap of 100 bp
* 48886 49960: contig of 1075 bp in length
* 49961 50060: gap of 100 bp
* 50061 51132: contig of 1072 bp in length

```

```

* 51133 51232: gap of 100 bp
* 51233 52356: contig of 1124 bp in length
* 52357 52456: gap of 100 bp
* 52457 53556: contig of 1100 bp in length
* 53557 53656: gap of 100 bp
* 53657 54782: contig of 1126 bp in length
* 54783 54882: gap of 100 bp
* 54883 56008: contig of 1126 bp in length
* 56009 56108: gap of 100 bp
* 56109 57307: contig of 1199 bp in length
* 57308 57407: gap of 100 bp
* 57408 58529: contig of 1122 bp in length
* 58530 58629: gap of 100 bp
* 58630 59799: contig of 1170 bp in length
* 59800 59899: gap of 100 bp
* 59900 60983: contig of 1084 bp in length
* 60984 61083: gap of 100 bp
* 61084 62193: contig of 1110 bp in length
* 62194 62293: gap of 100 bp
* 62294 63397: contig of 1104 bp in length
* 63398 63497: gap of 100 bp
* 63498 64574: contig of 1077 bp in length
* 64575 64674: gap of 100 bp
* 64675 65800: contig of 1126 bp in length
* 65801 65900: gap of 100 bp
* 65901 65993: contig of 1093 bp in length.
FEATURES
    source
    1. .66993
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="18"
    /map="18"
    /clone="RP13-907D8"
    /clone_lib="RPCI-13 Human Female BAC"

ORIGIN
Query Match      7.8%; Score 67.6; DB 2; Length 66993;
Best Local Similarity 33.5%; Pred. No. 0.0098;
Matches 278; Conservative 0; Mismatches 553; Indels 0; Gaps 0;

Qy 6 TTTCGTCATCACTTAAAGTTATGAGATATATCCATGTTGAATTTTGTAGCTGGT 65
Db 56412 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56471

Qy 66 TGATGATTTTACATATATAGTATATACATCCATGATAGTCTCAGTAGATAATCCTC 125
Db 56472 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56531

Qy 126 CTATGTTTAAACATTCGCTGCTCTCATTTGACCTATTTTAAACAGCCCTCTTAAA 185
Db 56532 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56591

Qy 186 TATACATCTCTGCTAGTCTGCTAGATAAGAGTGGCTGAATAAAGTGGCTGAATC 245
Db 56592 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56651

Qy 246 ATCTTCAACTCTAGTAGATGTCACAACTGTTTCTAAAGTGTATTTAACTATTAT 305
Db 56652 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56711

Qy 306 ATGTCATTTTGAACAGCTCTTCACTTACTAGCAATTTATATACAGACACTGTTAT 365
Db 56712 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56771

Qy 366 TGTACAGACTTTAAAGTTTTCATTCAGCTGGNNNNNNNNNNNNNNNNNNNNNNNN 425
Db 56772 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56831

Qy 426 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 485
Db 56832 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56891

```

```

Qy 486 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 545
Db 56892 TTTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTTNNNTT 56951

Qy 546 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 605
Db 56952 NNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 57011

Qy 606 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 665
Db 57012 TTTTNNNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 57071

Qy 666 GTTAAATATTTTCAAAATCGACATATGATATATATCTTCCAAAATTTTAAATTTTG 725
Db 57072 TTTNNNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 57131

Qy 726 TCTTTTTCACATTTTACTCTTTAGCTCAGCTGGAATTCATTTCTGTGTGTGAGAG 785
Db 57132 TTTTNNNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 57191

Qy 786 TAAGTCCTTTTTCATGTTTTCCTATGAATAAATAATTCCTCTCTGATT 836
Db 57192 TTTTNNNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 57242

RESULT 9
BX842689/c 146871 bp DNA linear HTG 03-DEC-2003
LOCUS Danio rerio clone DKEY-96D22, WORKING DRAFT SEQUENCE, 11 unordered
DEFINITION pieces.
ACCESSION BX842689
VERSION BX842689.1 GI:38677860
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 146871)
Burton, J.
Direct Submission
Submitted (02-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zk96D22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 14382 bases at least Q40
Consensus quality: 144621 bases at least Q30
Consensus quality: 145107 bases at least Q20
Insert size: 145871; sum-of-contigs
Insert size: 162937; 4.9% error; agarose-fp
Quality coverage: 5.37% in Q20 bases; sum-of-contigs Quality
coverage: 4.81% in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 19996: contig of 19996 bp in length
* * 19997 20096: gap of 100 bp
* * 20097 44027: contig of 23931 bp in length

```



```

QY 81 TATATAGTATACATT-CCATGGATAGTTCACAGTAGATAATCTCCTCTGTTTAAACAT 139
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3451 TATTTGATATATGTTGAATGGATATGGGAGTGTAGATTTTTTTTGGATATAGTATTT 3510
QY 140 TTGGTGCTCTCATTTTGACCTATTTTAAACAGGCTCTTTAAAFATACATCTCTG 199
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3511 TAAATTTTGGAGATATATTTAGAGTGGGATTTGGGTATATAGTAAATTTATGTT 3570
QY 200 TAGTGTATGCTAGAAATGGAGTGGCTGAATAAAGTGGCTGAATCATCTCAACTCTAG 259
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3571 TAGTGTATGGAGATCGTTATATATTTTAAATGGTGTATTAATTTATATTTTAA 3630
QY 260 TAAGATGCAACGTTTTCTAAAGTGTATTTTAACTATATATATGTCAAATTTTGA 319
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3631 TTAATGGTTTAAAGGTTTTTTTTTTTATATATTTTATTA-ATACGTATTATTTGTT 3668
QY 320 CAGCTCTTCACTTACTAGCAATTTATATCAGCAACACTGTATATGTCAGACTTTAA 379
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3689 TATTTGTTAAAGTTATTTAATAGATGATATTTTATTAATTTTAAATTTGATTTT 3748
QY 380 GTTTTCATTCACCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3749 TTGATAATTAGTGATGCGAGTGTGTTTTATATATTTTGTGGTTAGTTTATGTTT 3808
QY 440 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 499
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3809 TTTTAAATTCAGATGTTTCTTTCGTTTTATTGAGTGTGTAATTTTATATTTTGG 3868
QY 500 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 559
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3869 GATATTAGTTTTTATAGATATATGTTGTAAATATTTTTTTTAGTTTTTGGATAT 3928
QY 560 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 619
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3929 TTTTTCGTTTTTTTGTGTAGAAGTTTTTGTAGTTGATGTAATTTTTTGGTTATTT 3988
QY 620 NNNNNNNNGATCATATTTATGTTTGTCTATTTAAATAAAGCTGTTAAATTTTCA 679
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3989 TGAATTTGTTGTTTGTGAGTGTGTTGTGTAATAAATTTGTTAGAAAATGTT 4048
QY 680 AAATCGACATATGATATATTTCTCCAAAATTTTAAATAATTTGCTTTTTTTCACAT 739
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4049 ATAAAGATTTTTTTTTATGTTTTTTTTTTAGTATTTAATTTTCGGTTTTT---ATG 4105
QY 740 TTAGCTTTAGCTCAGCTGAATTCATTTCTGTGCTGTGTTGAGATAGTCTTTTTCAT 799
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4106 TAAGTTTTTAATTTATTTGAGTGTAGTTTTGTGTATAGTGTGGGATAAGTGTAAAT 4165
QY 800 GTTTTT 805
Db |||||
4166 TATTTT 4171

RESULT 11
BX548071
LOCUS
DEFINITION
  Danio rerio clone DKEY-25L23, WORKING DRAFT SEQUENCE, 4 unordered
  pieces.
ACCESSION
  BX548071.3
VERSION
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
  Danio rerio (zebrafish)
SOURCE
  Danio rerio
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Danio.
  1 (bases 1 to 239339)
REFERENCE
  McLaren,S.
  Direct Submission
  Submitted (24-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
  zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
  On Sep 24, 2003 this sequence version replaced gi:32479716.
COMMENT

```

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZK25L23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 237186 bases at least Q40
Consensus quality: 237537 bases at least Q30
Consensus quality: 237961 bases at least Q20
Insert size: 239039; sum-of-contigs
Insert size: 212898; 8.9% error; agarose-fp
Quality coverage: 9.82x in Q20 bases; sum-of-contigs Quality
coverage: 11.38x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 98607: contig of 98607 bp in length
* 98608 98707: gap of 100 bp
* 98708 142610: contig of 43903 bp in length
* 142611 142710: gap of 100 bp
* 142711 190543: contig of 47833 bp in length
* 190544 190643: gap of 100 bp
* 190644 239339: contig of 48696 bp in length.

```

FEATURES

```

source
  1..239339
    /organism="Danio rerio"
    /mol_type="genomic DNA"
    /db_xref="taxon:7955"
    /clone="DKEY-25L23"
    /clone_lib="DanioKey"
  1..98607
    /note="assembly_fragment:01005"
    clone_end.SP6
    vector_side:left
  98708..142610
    /note="assembly_fragment:01799"
    fragment_chain:1
  142711..190543
    /note="assembly_fragment:02192"
    fragment_chain:1
  190644..239339
    /note="assembly_fragment:02561"
    fragment_chain:1

```

ORIGIN

```

Query Match      7.68; Score 65.6; DB 2; Length 239339;
Best Local Similarity 31.5%; Pred.No. 0.016;
Matches 245; Conservative 0; Mismatches 532; Indels 0; Gaps 0;

QY 49 ATTTGTAGCTGCTGTTGATGATTTTACATATATAGTATTTACATTCATGATGTT 108
Db 23118 ATTTTATTTAGATTTTGTGTTTATTTAGATTTTGTGTTATATATATTTACTGTT 23177
QY 109 CTCAGTAGATAATCCCTCATTTGTTTAAACATTTGCGTTCTCTCTCATTTGACCTATTTT 168
Db 23178 TTATTTGTATTTATTTTATTTTATTTTCTTTTACTTTTCTTACTTTTGT 23237
QY 169 AAACAGCGCTCTTAAATATACACTTCTCTGTAGTGTAGTGTAGAAATGAGTGGCTGGA 228
Db 23238 ATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 23297
QY 229 ATAAAGTGCTGGAATCACTTCACTCACTAGTAGTAGTCAAACTGTTTCTTAAAGTGT 288

```


OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 2958.17 Seconds
(without alignments)
8711.839 Million cell updates/sec

Title: US-10-082-830-98

Perfect score: 963

Sequence: 1 atgggttttggtcattcaact.....gtgagaccaatgatagctg 863

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91.2	10.6	986	12	BG116503 602317590
2	89.8	10.4	679	13	BX488216 DKFZp686P
3	78	9.0	1200	13	BX437758 BX437758
4	78	9.0	1392	29	CG757503 P052-4-CO

5	72.6	8.4	1811	29	CG753732
6	71.2	8.3	1288	29	CG744915
7	69.8	8.1	1221	28	CC301561
8	69.2	8.0	1380	29	CG744815
9	68.2	8.0	1389	29	CG747695
10	68.6	7.9	1531	29	CG748014
11	68.4	7.9	1297	29	CG758143
12	68.2	7.9	1491	29	CG753221
13	68.2	7.8	1377	29	CG749971
14	66.8	7.7	1376	29	CG747831
15	66.6	7.7	1276	29	CG754010
16	66.6	7.7	1353	29	CG74812
17	66	7.6	1566	29	CG757757
18	65.6	7.6	1135	29	CNS033GQ
19	65.6	7.6	1434	28	CC187638
20	65.2	7.6	1361	29	CG744327
21	65	7.5	973	29	CNS0711E
22	65	7.5	1139	28	AQ897537
23	65	7.5	1313	29	CG751144
24	65	7.5	1352	28	CC288594
25	64.8	7.5	1225	29	CG745927
26	64.8	7.5	1331	29	CG749578
27	64.6	7.5	1260	29	CG751608
28	64.6	7.5	1283	29	CG749744
29	64.4	7.5	1373	29	CG750869
30	64.4	7.5	1433	29	CG745119
31	64.2	7.4	1101	29	CNS003BD
32	64.2	7.4	1298	29	CG756607
33	64.2	7.4	1362	29	CG757918
34	64.2	7.4	1528	29	CG753854
35	64	7.4	1324	29	CG746828
36	63.8	7.4	1254	29	CG748584
37	63.8	7.4	1626	14	CF238805
38	63.6	7.4	980	29	CNS0141E
39	63.4	7.3	1224	29	CG757393
40	63.4	7.3	1235	29	CG750878
41	63.4	7.3	1269	29	CG757211
42	63.4	7.3	1426	28	CC231597
43	63.2	7.3	1018	12	BM416001
44	63.2	7.3	1063	13	BX414736
45	63	7.3	1211	29	CG747324

ALIGNMENTS

RESULT 1
BG116503
LOCUS 602317590F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4417870 5',
DEFINITION mRNA sequence.
ACCESSION BG116503.1 GI:12610009
VERSION BG116503
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 986)
AUTHORS NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10151 row: c column: 23
High quality sequence stop: 634.

CG753732 P048-4-GO
CG744915 P037-3-F0
CC301561 CH261-13K
CG744815 P037-3-B0
CG747695 P041-2-D0
CG748014 P041-4-B0
CG758143 P053-3-B1
CG753221 P048-2-A0
CG749971 P044-3-E0
CG747831 P041-3-B0
CG754010 P049-3-C0
CG74812 P037-3-B0
CG757757 P053-1-D0
AL226115 Tetradon
CC187638 CH261-98P
CG744327 P036-4-E0
AL425064 clone BAO
AQ897537 HS_3153_A
CG751144 P045-1-E0
CC288594 CH261-67F
CG745927 P039-1-A0
CG749578 P043-4-D1
CG751608 P046-1-C1
CG749744 P044-1-C1
CG750869 P045-2-E1
CG745119 P037-4-G0
AL064091 Drosophil
CG756607 P051-4-C0
CG757918 P053-2-A0
CG753854 P049-1-D0
CG746828 P040-1-G1
CG748584 P042-3-B0
CF238805 AGENCOURT
AL103580 Drosophil
CG757393 P052-3-G0
CG750878 P045-2-F0
CG757211 P052-2-G0
CC231597 CH261-36A
BM416001 OP21084_M
BX414736 BX414736
CG747324 P040-4-D1

BG116503 986 bp mRNA linear EST 30-JAN-2001
602317590F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4417870 5',
mRNA sequence.

BG116503.1 GI:12610009
BG116503
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 986)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM10151 row: c column: 23
High quality sequence stop: 634.

```

FEATURES
  source
    Location/Qualifiers
      1..986
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:441780"
        /tissue_type="duodenal adenocarcinoma, cell line"
        /lab_hosts="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_88"
        /note="Organ: small intestine; Vector: pCMV-SPORT6;
        Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
        oligo-dT primed. Average insert size 1.767 kb. Library
        enriched for full-length clones and constructed by Life
        Technologies. Note: this is a NIH_MGC library."

ORIGIN
  Query Match      10.6%; Score 91.2; DB 12; Length 986;
  Best Local Similarity 96.9%; Pred. No. 5.7e-08;
  Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 166 TTTAAACAGGCGCTCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCT 225
Db 336 TCTGAAGAGCGCTCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCT 395

Qy 226 GGAATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 396 GGATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 431

RESULT 2
EX488216
LOCUS
DEFINITION
  EX488216 Homo sapiens 679 bp mRNA linear EST 04-SEP-2003
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ACCESSION
  EX488216
VERSION
  EX488216.1 GI:31953581
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
REFERENCE
  1 (bases 1 to 679)
  Bahr, A., Lauber, J., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
  Fobo, G., Han, M. and Wiemann, S.
  EST (Bahr, A., Lauber, J., Mewes, H.W., Weil, B., et al.)
  Unpublished (2003)
  Contact: MIPS
  MIPS
  Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
  This is the 5' sequence of the clone insert
  Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
  Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;
  sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
  consortium of the German Genome Project.
  No s1 sequence available.
  This clone (DKFZp686p11268) is available at the RZPD in Berlin.
  Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
  Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
  Location/Qualifiers
    1..679
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="DKFZp686p11268"
      /dev_stage="adult"
      /lab_host="DH10B"
      /clone_lib="686 (synonym: hlcc3)"
      /notes="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
      cDNA-collection"

FEATURES
  source
    Location/Qualifiers
      1..1200
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0CAP008YB01"
        /tissue_type="THYMUS"
        /clone_lib="Homo sapiens THYMUS"
        /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
        with a NotI-oligo(dT) primer. Five prime end enriched,
        double-strand cDNA was digested with Not I and cloned into
        the Not I and EcoRV sites of the pCMVSPORT 6 vector.
        Library was not normalized."

FEATURES
  source
    Location/Qualifiers
      1..1200
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0CAP008YB01"
        /tissue_type="THYMUS"
        /clone_lib="Homo sapiens THYMUS"
        /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
        with a NotI-oligo(dT) primer. Five prime end enriched,
        double-strand cDNA was digested with Not I and cloned into
        the Not I and EcoRV sites of the pCMVSPORT 6 vector.
        Library was not normalized."

ORIGIN
  Query Match      9.0%; Score 78; DB 13; Length 1200;
  Best Local Similarity 24.9%; Pred. No. 1.8e-05;
  Matches 200; Conservative 95; Mismatches 505; Indels 3; Gaps 1;

Qy 20 TTTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGTGTTGATGATTTTACA 79
Db 979 WYATATAAAAAAAWATTTTWTCTTTTCTTTTWTTTTWTTTTWTAAATTTTTT 920

Qy 80 TTATATAGTATTTACATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 139
Db 919 TTATAAAWTTTTTTTATTTTTTTTOMWYVTVTTTTTATTTTATTTTATTTATATA 860

Qy 140 TTGGTTGCTTCTCATTTTGAACCTATTTTAAACAGGCTCTTTAAATATACACTTCTCT 199
Db 859 AAWATAATATTWYTYWTTTTTTTTTTTWTWATWATNTTCTTCTTCTTCTTCTTCTTCT 800

Qy 200 TAGGT--ATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCTGATCATCTTCAACTC 256
Db 799 WTTTTTMTATCTTWTWAAAAAATWATATAAAAAAHTCTCTYATATAWTTAAWATC 740

Qy 257 TAGTAAGATGTCAACTGTTTCTTCAAAAGTGTGTTTATTTAACTATTTATGTCATTTT 316
Db 739 YTTCTMAAAAAAAWTTTCTCTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 680

```


[illegible][illegible]

```

Class: BAC ends
High quality sequence start: 73
High quality sequence stop: 274.
Location/Qualifiers
    1. 1221
        /organism="Gallus gallus"
        /mol_type="genomic DNA"
        /strain="Red Jungle Fowl"
        /db_xref="taxon:9031"
        /clone="CH261-13K20"
        /sex="female"
        /cell_line="UCD001, inbred 256"
        /clone_lib="CH261"
        /notes="vector: pTABBAC3.1; Site.1: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: http://www.chori.org/bacpac"

```

ORIGIN	Query Match	8.1%	Score 59.8	DB 28	Length 1221
	Best local Similarity	29.9%	Pred. No. 0.00063		
	Matches 240	Conservative 0	Mismatches 564	Indels 0	Gaps 0
QY	2	TGGGTTTGGGTCATCAACTTAAAGTTTATGAGATATATCCATGTGTAATTTGTAGCTGT	61		
DB	405	TNNNTTTTNNNNNNNTTTTTTNNNNNTTTTNTTTTTTNTTTTTTTTTTTTTTTTTTTT	464		
QY	62	GGTTTGAATGATTTTACATATATAGTATTACATTCCTCATGGATAGTTCTCAGTAGATAAT	121		
DB	465	TTTTTTNNNTTTNTTTTTTTNTTTTTTTTTNNNTNNNTNTTTTTNNNTNTNTNTNTN	524		
QY	122	CCTCCTATGTTTAAACATTTGCGTGTCTCATTTTGACCTATTTTAAACAGCGCTCT	181		
DB	525	NNNTTTTNTTNTTTTTTTTTTNTTTTTTNTTTTTTNTTTTTTNNNNNTTNTTTTTT	584		
QY	182	TAAATATACACTTCTCTGTAGTGATGCTAGAAATGGAGTGGCTGGGAATAAAAGTGGCTG	241		
DB	585	TNTTTTTTNNNTTTNNNTNNNTTTTTTTTTTNTTTTTTNTTTTTTNTTTTTTNTTTT	644		
QY	242	AATCATCTTCAACTCTAGTAAGATGTCAAACGTGTTTCTAAAAGTGTTTATATTTAACTA	301		
DB	645	TTTNTTNTTTTTTCCCTTTTNTTNTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTN	704		
QY	302	TTATATGTCAAATTTTGAACACTCTTTACACTTACTAGCAATTTATATCAGCAACACTG	361		
DB	705	TNTTTTTTNTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTNTTNTTNN	764		
QY	362	TTATGTGCAGACTTTTAAAGTTTTCATTCACCTGGNNNNNNNNNNNNNNNNNNNNNN	421		
DB	765	TNNNTTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTNTTNTTNTTNTTTTTTTTTTTT	824		
QY	422	NN	481		
DB	825	NNTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	884		
QY	482	NN	541		
DB	885	TTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNT	944		
QY	542	NN	601		
DB	945	TTTTTTTTTNTTTTTTTNTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1004		
QY	602	NNNNNNNNNNNNNNNNNNNNNNNNNGTGACATATTATGTTTGTCTATTATAAAAAA	661		
DB	1005	TNNTTTTTTNTTTTTTTTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTT	1064		
QY	662	CTGCTTAAATATTTCAAAATTCGACATATGATATATATCTTCCAAAATTTTAAATAT	721		
DB	1065	TTTTTTTTTNTTNTTNTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	1124		
QY	722	TTTGCTTTTTTTCACATTTTAGTCTTTTASCTCAGCTGGAAATCATTTCTGTGTGGTGT	781		
DB	1125	TT	1184		

```

QY 782 GAGATAAGTCCTTTTTCATGTTTT 805
Db 1185 TTTTNNNNNTNNNTTTTTTTTT 1208

RESULT 8
CG744815 1380 bp DNA linear GSS 24-OCT-2003
LOCUS P037-3-B04.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION genomic survey sequence.
ACCESSION CG744815
VERSION CG744815.1 GI:37965683
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1380)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1. .1380
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="Californio"
/db xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/notes="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

FEATURES
source
Query Match 8.0%; Score 69.2; DB 29; Length 1380;
Best Local Similarity 32.4%; Pred. No. 0.00087;
Matches 269; Conservative 0; Mismatches 562; Indels 0; Gaps 0;

QY 6 TTGGTCATTCACCTTAAGTTATGAGATATACCATGTCATGTTGAGTTTGTAGCTGTGGTT 65
Db 527 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 586

QY 66 TGATGATTTTACATATATAGTATTACATTCATGATAGTTCACAGTAGATAATCCTC 125
Db 587 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 646

QY 126 CTATGTTTAAACATTTGCGTGTCTCTCATTTTGACCTATTTTAAACAGGCCTCTTTAAA 185
Db 647 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 706

QY 186 TATACACTTCTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 245
Db 707 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 766

QY 246 ATCTCAACTCTAGTAAGATGTCACAACTGTTTCTAAAGTGTATTATTAACTATTAT 305
Db 767 TTTTCCCTTTTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 826

QY 306 ATGCAATTTTGAAGCTCTTCACTTACTAGCAATTTATTATCAGCAACACTTGTAT 365
Db 827 NNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 886

```

```

QY 365 TCTCAGACTTTTACGTTTTCATTCACGTCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 425
Db 887 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 946

QY 426 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 485
Db 947 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 1006

QY 486 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 545
Db 1007 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 1066

QY 546 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 605
Db 1067 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 1126

QY 606 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 665
Db 1127 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 1186

QY 666 GTTAATAATTCACAAATCGACATTATGATAATATATCTTCCAAAATTTTAATAATTTTG 725
Db 1187 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 1246

QY 726 TCTTTTTCACATTTTAGCTTTAGCTCAGCTGGAATTCATTTCTGTGTGGTGTGAGA 785
Db 1247 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1306

QY 786 TAAGTCTTTTTCATGTTTTTCCCTATGAAATAAATAATTTCCCTCTCTATT 836
Db 1307 TTTTNNNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNTNNNT 1357

```

```

RESULT 9
CG747695 1389 bp DNA linear GSS 24-OCT-2003
LOCUS P041-2-D06.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION genomic survey sequence.
ACCESSION CG747695
VERSION CG747695.1 GI:37968621
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

```

```

REFERENCE 1 (bases 1 to 1389)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1. .1389
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="Californio"
/db xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/notes="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

```

```

FEATURES
source

```

```

ORIGIN

```



```

QY 606 NNNNNNNNNNNNNNNNNNNNGTGATCATATTATGTTTGTCTCAATTAATAAAACCTG 665
Db 720 TTNNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 661
QY 666 GTTAAATATTTACAAATCGACATATGATATATCTTCCAAATTTTAAATAATTTG 725
Db 660 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 601
QY 726 TCTTTTTTTCACATTTTCTAGCTCAGCTGAGATTCATCTCTCTCTGTTAT 785
Db 600 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 541
QY 786 TAAGTCCTTTTTCATGTTTTCCTCATGAAATATAATTCCTCTCTGTTAT 836
Db 540 TTNNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 490

RESULT 11
CG758143
LOCUS
DEFINITION
P053-3-B11.2b Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION
CG758143
VERSION
CG758143.1 GI:37987385
KEYWORDS
GSS.
SOURCE
Pristionchus pacificus
ORGANISM
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 1297)
REFERENCE
AUTHORS
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE
An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL
Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE
22835951
PUBMED
12884007
COMMENT
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1..1297
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/notes="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

FEATURES
source
Query March 7.9%; Score 68.4; DB 29; Length 1297;
Best Local Similarity 32.1%; Pred. No. 0.0013;
Matches 258; Conservative 0; Mismatches 546; Indels 0; Gaps 0;

ORIGIN
QY 6 TTTGGTCATTCAACTTTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGGTT 65
Db 492 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 551
QY 66 TGATGATTTTACATATATAGTATTACATTCATCGATAGTTCAGTAGATATCCCTC 125
Db 552 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 611
QY 126 CTATTGTTTAAACATTTGCGTGTCTCTCATTTTGTACCTATTTTAAACGGCTCTTAAA 185
Db 612 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 671

```

```

QY 186 TATACACTCTCTCTAGTGTATGCTAGAAATGGAGTGGCTGAATAAAGTGGCTGAATC 245
Db 672 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 731
QY 246 ATCTTCAACTCTAGTAAGATGTCAAACTGTTTCTTAAAGTGTGTTTTTAACTATTAT 305
Db 732 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 791
QY 306 ATGTCAATTTTGAACAGCTCTTCTACTAGCAATTTATATCAGCAACACTTGTAT 365
Db 792 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 851
QY 366 TGTCAAGACTTTTAACTTTTTCATCTCACTCGNNNNNNNNNNNNNNNNNNNNNNNN 425
Db 852 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 911
QY 426 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 485
Db 912 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 971
QY 486 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 545
Db 972 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1031
QY 546 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 605
Db 1032 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1091
QY 606 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 665
Db 1092 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1151
QY 666 GTTAAATATTTTCAAAAACGACATATGATATATATCTTCCAAATTTTAAATAATTTG 725
Db 1152 NTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1211
QY 726 TCTTTTTCACATTTTGTAGCTTGTAGCTCAGCTGGAATTCATCTGTGTGTGTGAGA 785
Db 1212 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1271
QY 786 TAAGTCCTTTTTCATGTTTTCCTCCT 809
Db 1272 TTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 1295

RESULT 12
CG753221
LOCUS
DEFINITION
P048-2-A01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION
CG753221
VERSION
CG753221.1 GI:37977480
KEYWORDS
GSS.
SOURCE
Pristionchus pacificus
ORGANISM
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 1491)
REFERENCE
AUTHORS
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE
An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL
Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE
22835951
PUBMED
12884007
COMMENT
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

```


[illegible]

RESULT 14
CG747831
LOCUS CG747831 1376 bp DNA linear GSS 24-OCT-2003
DEFINITION P041-3-B05.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.

VERSION	CG4747831.1	GI:37968757
KEYWORDS	GSS.	
SOURCE	Pristionchus pacificus	
ORGANISM	Pristionchus pacificus	
	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.	
REFERENCE	1 (bases 1 to 1376)	
AUTHORS	Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Bunjer,J., van der Meulen,M. and Sommer,R.J.	
TITLE	An integrated physical and genetic map of the nematode Pristionchus pacificus	
JOURNAL	Mol. Genet. Genomics	269 (5), 715-722 (2003)
MEDLINE	22835951	
PUBLISHED	12884007	

Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de

[illegible]

RESULT 15
 CG754010
 LOCUS
 DEFINITION
 CG754010 1276 bp DNA linear GSS 24-OCT-2003
 P049-2-C03.ya Ppa EcoRI BAC Library *Pristionchus pacificus* genomic,
 genomic survey sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 CG754010
 CG754010.1 GI:37979071
 GSS.
Pristionchus pacificus
Pristionchus pacificus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasteridas; *Pristionchus*.
 1 (bases 1 to 1276)
 Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
 Bunjer,J., van der Meulen,M. and Sommer,R.J.
 An integrated physical and genetic map of the nematode *Pristionchus*
pacificus

JOURNAL
MEDLINE
PUBMED
COMMENT

Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de

FEATURES
SOURCE

ORIGIN

Query Match		7.7%	Score 66.6	DB 29	Length 1276
Best Local Similarity		30.0%	Pred. No. 0.0028		
Matches 249		Conservative 0	Mismatches 582	Indels 0	Gaps 0
QY	6	TTTGGTCATTCAACTTTAAGTTTATGAGATATCCATGTTGAATTGTAGCTGTGGTT	65		
DB	256	TTTTTTTNTTT	315		
QY	66	TGATGATTTTTACATTATATAGTATTACATTCCTGGATAGTCTCAGTAGATAATCCTC	125		
DB	316	TTNTNNNTT	375		
QY	126	CTATTGTTTAACTTTGGGTGCTCTCAATTTGACCTATTATTTAAACAGGCCTCTTAAA	185		
DB	376	TTT	435		
QY	196	TATACACTCTCTGTFAGTGTATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCTGAATC	245		
DB	436	TTNTT	495		
QY	246	ATCTTCAACTCTPAGTAAAGATGTCAAACGTGTTCTTCAAAGTGTGTTTATTTTAACTATTAT	305		
DB	496	NTTT	555		
QY	306	ATGTCAAATTTGAAACAGCTCTTCTCACTTACIAGCAATTTATATCAGCAACACTTGTAT	365		
DB	556	TTTNTTNT	615		
QY	366	TGTCAGACTTTTAAAGTTTTCATTCCTCGNNNNNNNNNNNNNNNNNNNNNNNNNNNN	425		
DB	616	TTT	675		
QY	426	NN	485		
DB	676	TTT	735		
QY	486	NN	545		
DB	736	TTT	795		
QY	546	NN	605		
DB	796	TTT	855		
QY	606	NNNNNNNNNNNNNNNNNNNNNNNGTGATCATATTATGTTTGTCTCATTTTAAAAAACCTG	665		
DB	856	TTT	915		
QY	666	GTTTAAATATTTCACAAATCGACATTATGATATATTATCTTCCAAAAATTTTAAATTTTG	725		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 434.526 Seconds
(without alignments)
8437.229 Million cell updates/sec

Title: US-10-082-830-98
Perfect score: 863
Sequence: 1 atgggtttgtcattcaact.....gttgagaccatgatagctg 863

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002s:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	630	73.0	863	6	ABK93492 Human bre
2	392	45.4	392	6	ABK93491 Human bre
3	89.8	10.4	229	4	AAL26040 Human bre
C 4	89.8	10.4	233	4	AAL16838 Human bre
C 5	89.8	10.4	256	4	AAL17847 Human bre
6	89.8	10.4	258	4	AAL17770 Human bre
7	89.8	10.4	270	4	AAL10695 Human bre
C 8	89.8	10.4	298	4	AAL10772 Human bre
9	89.8	10.4	317	4	AAL17429 Human bre
10	89.8	10.4	332	4	AAL09537 Human bre
C 11	85	9.8	525	4	AAL21979 Human bre
C 12	77.8	9.0	248	4	AAL08295 Human bre
C 13	71.6	8.3	534	4	AAL13109 Human bre
14	67.6	7.8	6062	6	AAS61094 Human gen
15	66.2	7.7	6195	6	ABL32591 Human imm
16	65.4	7.6	3683	7	ABZ10199 Haematopo
17	62.8	7.3	7128	6	ABL33559 Human imm
18	62.8	7.3	40324	6	ABQ67150 Human ang
19	62.2	7.2	11147	7	AZL10154 Haematopo
20	62.2	7.2	14147	9	ABZ54226 Pretrate
21	62.2	7.2	14147	9	ABE84164 Human lym
22	60.6	7.0	9964	6	ABL32098 Human imm
23	60.4	7.0	11147	7	ABZ10008 Haematopo

24	60.4	7.0	14147	4	AAS46744 Tumour su
25	60.4	7.0	14147	6	ABK33956 Human DNA
26	60.4	7.0	14147	7	ADA20387 Prostate
27	60.4	7.0	14147	7	ADA84194 Human ren
28	60.4	7.0	14147	9	ADB54098 Precrete
29	60.4	7.0	14147	9	ADE84088 Human lym
30	59.6	6.9	7442	4	AAS46686 Tumour su
31	59.6	6.9	15732	4	AAS45388 Chemicall
32	59.6	6.9	15732	6	ABK28233 DNA trans
33	59.4	6.9	6171	6	ABL33010 Human imm
34	59.2	6.9	3683	7	ABZ10053 Haematopo
35	58.8	6.8	5327	4	AAS45356 Chemicall
36	58.8	6.8	5327	6	ABK28189 DNA trans
37	58.8	6.8	9539	4	AAS45347 Chemicall
38	58.8	6.8	12237	6	ABK28180 DNA trans
39	58.6	6.8	12237	6	ABL34358 Human imm
40	58.2	6.7	6132	6	ABL32863 Human imm
41	58	6.7	12356	4	AAS46510 Tumour su
42	57.6	6.7	6000	6	ABL34305 Human imm
43	57.6	6.7	6292	4	AAS46735 Tumour su
44	57.6	6.7	10467	6	ABL49302 Human pol
45	57.2	6.6	5362	6	ABK28413 DNA trans

ALIGNMENTS

RESULT 1
ABK93492
ID ABK93492 standard; CDNA; 863 BP.
XX
AC ABK93492;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human breast specific nucleic acid, BSNA #98.
XX
KW Human; ss; breast specific nucleic acid; BSNA; breast cancer;
KW mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.
XX
OS Homo sapiens.
XX
PN WO200236807-A2.
XX
PD 10-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-US046888.
XX
PR 27-OCT-2000; 2000US-0243802P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
XX
DR WPI; 2002-463415/49.
XX
FT New breast-specific nucleic acids and polypeptides, useful for
FT identifying, diagnosing, monitoring, staging, imaging, and treating
FT breast cancer and non-cancerous disease states in breast tissues.
XX
PS Claim 1; Page 191; 281pp; English.
XX
CC The invention relates to breast specific nucleic acids (BSNA) and breast-
CC specific polypeptides (BSP). Also included are a method for determining
CC the BSNA in a sample, a vector comprising a BSNA, a host cell comprising
CC the vector, a method for producing a polypeptide encoded by a BSNA, an
CC anti-BSP antibody and a method for determining the presence of a BSP in a
CC sample. The breast-specific nucleic acids, polypeptides and compositions
CC comprising them are useful for identifying, diagnosing, monitoring and
CC staging, imaging, and treating breast cancer, mammary tumour and non-
CC cancerous disease states in breast tissue; for identifying breast tissue;
CC for monitoring, identifying and/or designing agonists and antagonists of
CC the polypeptides; in gene therapy; in producing transgenic animals and

CC cells; for producing engineered breast tissue for treatment and research;
 CC and as elements in an array or computer program for pattern recognition
 CC of breast disorders. The nucleic acids may be used as hybridisation
 CC probes to detect, characterise and quantify hybridising nucleic acids in,
 CC and isolate hybridising nucleic acids from, both genomic and transcript-
 CC derived nucleic acid samples. The BSP protein may be used in a vaccine
 CC composition for raising an immune response against breast cancer. The
 CC present sequence is BSNA cDNA of the invention
 XX
 SQ Sequence 863 BP; 173 A; 94 C; 98 G; 265 T; 0 U; 233 Other;

Query Match 73.0%; Score 630; DB 6; Length 863;
 Best Local Similarity 100.0%; Pred. No. 1.8e-121;
 Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTGGTCATTCACCTTTAAGTTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 DB 1 ATGGGTTGGTCATTCACCTTTAAGTTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 QY 61 TGGTTGATGATTTTACATTATATAGTATATACATTCATGATGATTTTCAAGTATATA 120
 DB 61 TGGTTGATGATTTTACATTATATAGTATATACATTCATGATGATTTTCAAGTATATA 120
 QY 121 TCCTCCCTATGTTTAAACATTTGCGTCTCTCTCATTTTGACCTATTTTAAACAGGCTCT 180
 DB 121 TCCTCCCTATGTTTAAACATTTGCGTCTCTCTCATTTTGACCTATTTTAAACAGGCTCT 180
 QY 181 TTAATATACACTTCTCTAGTATGCTAGAAATGGAGTGGCTGGAATAAAGTGCT 240
 DB 181 TTAATATACACTTCTCTAGTATGCTAGAAATGGAGTGGCTGGAATAAAGTGCT 240
 QY 241 GAATCATCTCACTCTAGTAAAGTCAACTGTTTCTAAAGTGTATTTTAACT 300
 DB 241 GAATCATCTCACTCTAGTAAAGTCAACTGTTTCTAAAGTGTATTTTAACT 300
 QY 301 ATTATATGTCATTTTGAACAGCTCTTCTACTAGCAATTTATATCAGCAACACTT 360
 DB 301 ATTATATGTCATTTTGAACAGCTCTTCTACTAGCAATTTATATCAGCAACACTT 360
 QY 361 GTTATGTCAGCTTTTAAAGTTTCTACTGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
 DB 361 GTTATGTCAGCTTTTAAAGTTTCTACTGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
 QY 421 NNN 480
 DB 421 NNN 480
 QY 481 NNN 540
 DB 481 NNN 540
 QY 541 NNN 600
 DB 541 NNN 600
 QY 601 NNN 660
 DB 601 NNN 660
 QY 661 ACCTGGTTAAATATTTTACAAATCGCATATATATATATCTTCCAAATTTTAAATA 720
 DB 661 ACCTGGTTAAATATTTTACAAATCGCATATATATATATCTTCCAAATTTTAAATA 720
 QY 721 TTTTGTCTTTTTCACATTTTGTCTTTAGCTCAGCTGGAATTTCTTCTGTGTGGTG 780
 DB 721 TTTTGTCTTTTTCACATTTTGTCTTTAGCTCAGCTGGAATTTCTTCTGTGTGGTG 780
 QY 781 TGAGATAAGTCTTTTTCATGTTTTCCTTATGAATAAATAATTTCTTCTGTATTCAG 840
 DB 781 TGAGATAAGTCTTTTTCATGTTTTCCTTATGAATAAATAATTTCTTCTGTATTCAG 840
 QY 841 GTAGCTGAGACCAATGATAGCTG 863
 DB 841 GTAGCTGAGACCAATGATAGCTG 863

Db 841 GTAGCTGAGACCAATGATAGCTG 863
 RESULT 2
 ID ABK93491
 XX ABK93491 standard; cDNA; 392 BP.
 AC ABK93491;
 XX
 DT 23-AUG-2002 (first entry)
 XX
 DE Human breast specific nucleic acid, BSNA #97.
 XX
 KW Human; ss; breast specific nucleic acid; BSNA; breast cancer;
 KW mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200236807-A2.
 XX
 PD 10-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-US046888.
 XX
 PR 27-OCT-2000; 2000US-0243802P.
 XX
 PA (DIAD-) DIADEXUS INC.
 XX
 PI Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
 DR WPI; 2002-463415/49.
 XX
 PT New breast-specific nucleic acids and polypeptides, useful for
 PT identifying, diagnosing, monitoring, staging, imaging, and treating
 PT breast cancer and non-cancerous disease states in breast tissues.
 PS
 XX Claim 1; Page 191; 281pp; English.
 CC
 CC The invention relates to breast specific nucleic acids (BSNA) and breast-
 CC specific polypeptides (BSP). Also included are a method for determining
 CC the BSNA in a sample, a vector comprising a BSNA, a host cell comprising
 CC the vector, a method for producing a polypeptide encoded by a BSNA, an
 CC anti-BSP antibody and a method for determining the presence of a BSP in a
 CC sample. The breast-specific nucleic acids, polypeptides and compositions
 CC comprising them are useful for identifying, diagnosing, monitoring,
 CC staging, imaging, and treating breast cancer, mammary tumour and non-
 CC cancerous disease states in breast tissue; for identifying breast tissue;
 CC for monitoring, identifying and/or designing agonists and antagonists of
 CC the polypeptides; in gene therapy; in producing transgenic animals and
 CC cells; for producing engineered breast tissue for treatment and research;
 CC and as elements in an array or computer program for pattern recognition
 CC of breast disorders. The nucleic acids may be used as hybridisation
 CC probes to detect, characterise and quantify hybridising nucleic acids in,
 CC and isolate hybridising nucleic acids from, both genomic and transcript-
 CC derived nucleic acid samples. The BSP protein may be used in a vaccine
 CC composition for raising an immune response against breast cancer. The
 CC present sequence is BSNA cDNA of the invention
 XX
 SQ Sequence 392 BP; 108 A; 60 C; 61 G; 163 T; 0 U; 0 Other;

Query Match 45.4%; Score 392; DB 6; Length 392;
 Best Local Similarity 100.0%; Pred. No. 3.7e-72;
 Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGGGTTGGTCATTCACCTTTAAGTTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 DB 1 ATGGGTTGGTCATTCACCTTTAAGTTTATGAGATATATCCATGTTCAATTTTGTAGCTG 60
 QY 61 TGGTTGATGATTTTACATTATATAGTATATACATTCATGATGATTTTCAAGTATATA 120
 DB 61 TGGTTGATGATTTTACATTATATAGTATATACATTCATGATGATTTTCAAGTATATA 120
 QY 121 TCCTCCCTATGTTTAAACATTTGCGTCTCTCTCATTTTGACCTATTTTAAACAGGCTCT 180

```

Db 121 TCTCTCTATTGTTTAAACATTTGGGTTGCTTCTCTCAATTTGACCTATTTTAAACAGGCGCTCT 180
Qy 181 TTAATATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCT 240
Db 181 TTAATATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCT 240
Qy 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACTGTTTCTAAAAGTGTATTATTTTAACT 300
Db 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACTGTTTCTAAAAGTGTATTATTTTAACT 300
Qy 301 ATTATATGTCATTTTGAACAGCTCTTTCACTTACTAGCAATTTATTTATCAGCAACACTT 360
Db 301 ATTATATGTCATTTTGAACAGCTCTTTCACTTACTAGCAATTTATTTATCAGCAACACTT 360
Qy 361 GTTATTGTCAGACTTTTAACTTTTCACTTCACT 392
Db 361 GTTATTGTCAGACTTTTAACTTTTCACTTCACT 392

RESULT 3
AAL26040/c
ID AAL26040 standard; cDNA; 229 BP.
XX AC AAL26040;
XX XX
XX 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 18497.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PR 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX XX
XX WPI; 2001-451856/48.
XX DR
XX PT New peptide useful as a marker for the diagnosis of breast cancer.
XX PS Claim 1; Page 3412; 3695pp; English.
XX XX
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity
XX XX
XX SQ Sequence 229 BP; 60 A; 55 C; 51 G; 63 T; 0 U; 0 Other;
XX CC Query Match 10.4%; Score 89.8; DB 4; Length 229;
XX CC Best Local Similarity 97.8%; Pred. No. 1.6e-09;
XX CC Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

169 AAACAGGCGCTCTTTAAATATACACTTCTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
170 AAAATGGCGCTCTTTAAATATACACTTCTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 111
229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
110 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 78

RESULT 4
AAL16838/c
ID AAL16838 standard; cDNA; 233 BP.
XX AC AAL16838;
XX XX
XX 07-DEC-2001 (first entry)
XX DT Human breast cancer expressed polynucleotide 9295.
XX DE Human; breast cancer; cell marker; cytostatic; ss.
XX KW Homo sapiens.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PR 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX XX
XX WPI; 2001-451856/48.
XX DR
XX PT New peptide useful as a marker for the diagnosis of breast cancer.
XX PS Claim 1; Page 1670; 3695pp; English.
XX XX
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity
XX XX
XX SQ Sequence 233 BP; 61 A; 55 C; 53 G; 64 T; 0 U; 0 Other;
XX CC Query Match 10.4%; Score 89.8; DB 4; Length 233;
XX CC Best Local Similarity 97.8%; Pred. No. 1.6e-09;
XX CC Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

169 AAACAGGCGCTCTTTAAATATACACTTCTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
174 AAAATGGCGCTCTTTAAATATACACTTCTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 115
229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
114 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 82

```

```

RESULT 5
AAL17847/c
ID AAL17847 standard; cDNA; 256 BP.
XX
AC AAL17847;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 10304.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 1837; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 256 BP; 72 A; 56 C; 57 G; 71 T; 0 U; 0 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 256;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AACAGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
Db 174 AAAATGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 115
QY 229 ATAAAGTGCTGAATCATCTTCAACTCTAGTA 261
Db 114 ATAAAGTGCTGAATCATCTTCAACTCTAGTA 82
RESULT 6
AAL17770
ID AAL17770 standard; cDNA; 258 BP.
XX
AC AAL17770;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 10227.
XX

```

```

KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 1823; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 258 BP; 70 A; 59 C; 57 G; 72 T; 0 U; 0 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 258;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AACAGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 228
Db 85 AAAATGGCGCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGA 144
QY 229 ATAAAGTGCTGAATCATCTTCAACTCTAGTA 261
Db 145 ATAAAGTGCTGAATCATCTTCAACTCTAGTA 177
RESULT 7
AAL10695
ID AAL10695 standard; cDNA; 270 BP.
XX
AC AAL10695;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 3152.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX

```


PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
DR New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX Claim 1; Page 583; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 270 BP; 70 A; 61 C; 63 G; 73 T; 0 U; 3 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 270;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 169 AAACAGGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
Db 97 AAAATGGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 156
Oy 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 157 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 189
RESULT 8
AAL10772/c
ID AAL10772 standard; cDNA; 298 BP.
XX
XX AAL10772;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 3229.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US000798.
XX
XX 14-JAN-2000; 2000US-0176077P.
XX
XX 14-MAR-2000; 2000US-0189167P.
XX
XX 24-MAR-2000; 2000US-0192099P.
XX
XX 29-MAR-2000; 2000US-0193480P.
XX
XX 15-MAY-2000; 2000US-0205230P.
XX
XX 09-JUN-2000; 2000US-0211315P.
XX
XX 25-JUL-2000; 2000US-0220534P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX Claim 1; Page 596; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is
XX afflicted with breast cancer by examining the correlation between the
XX expression of certain markers and the cancerous state of breast cells.
XX The polynucleotides and encoded polypeptides are potential markers for
XX detecting, diagnosing, monitoring, characterising treating and
XX potentially preventing breast cancer. The polynucleotides and encoded
XX polypeptides are also useful for isolating compounds with cytostatic
XX activity
XX
SQ Sequence 298 BP; 77 A; 69 C; 73 G; 78 T; 0 U; 1 Other;
Query Match 10.4%; Score 89.8; DB 4; Length 298;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 169 AAACAGGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
Db 213 AAAATGGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 154
Oy 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
Db 153 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 121
RESULT 9
AAL17429
ID AAL17429 standard; cDNA; 317 BP.
XX
XX AAL17429;
XX
XX 07-DEC-2001 (first entry)
XX
XX Human breast cancer expressed polynucleotide 9886.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US000798.
XX
XX 14-JAN-2000; 2000US-0176077P.
XX
XX 14-MAR-2000; 2000US-0189167P.
XX
XX 24-MAR-2000; 2000US-0192099P.
XX
XX 29-MAR-2000; 2000US-0193480P.
XX
XX 15-MAY-2000; 2000US-0205230P.
XX
XX 09-JUN-2000; 2000US-0211315P.
XX
XX 25-JUL-2000; 2000US-0220534P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
XX
XX New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX Claim 1; Page 1762; 3695pp; English.
XX
XX The invention relates to human breast cancer expressed polynucleotides
XX (AAL07544-AAL26789) and methods of assessing whether a patient is

CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity

SQ Sequence 317 BP; 86 A; 69 C; 77 G; 85 T; 0 U; 0 Other;

Query Match 10.4%; Score 89.8; DB 4; Length 317;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 169 AAACAGGCGCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
DB 139 AAAATGGCGCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 198

QY 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 199 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 231

RESULT 10
ID AAL09537 standard; cDNA; 332 BP.
AC AAL09537;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 1994.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PR 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189157P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 389; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity

SQ Sequence 332 BP; 86 A; 75 C; 85 G; 86 T; 0 U; 0 Other;

Query Match 10.4%; Score 89.8; DB 4; Length 332;
Best Local Similarity 97.8%; Pred. No. 1.7e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 169 AAACAGGCGCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
DB 154 AAAATGGCGCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 213

QY 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 214 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 246

RESULT 11
ID AAL21979/c
XX AAL21979 standard; cDNA; 525 BP.
AC AAL21979;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 14436.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PR 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189157P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 2592; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity

SQ Sequence 525 BP; 127 A; 104 C; 110 G; 184 T; 0 U; 0 Other;

Query Match 9.8%; Score 85; DB 4; Length 525;
Best Local Similarity 94.6%; Pred. No. 1.8e-08;
Matches 88; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 169 AAACAGGCGCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
DB 387 AAAATGGCGCTTTAAATATACACTTCTCTAGTGTATGCTAGAAATGGAGTGGCTGGA 328

QY 229 ATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261

```

Db 327 ATAAAAGTGGCTGAATCATCTCAACTTAGTA 295
RESULT 12
AAL08295/c
ID AAL08295 standard; cDNA; 248 BP.
XX AC AAL08295;
XX DT 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 752.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX PF 10-JAN-2001; 2001WO-US000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX PF (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX DR New peptide useful as a marker for the diagnosis of breast cancer.
XX PT Claim 1; Page 208; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity
XX SQ Sequence 248 BP; 61 A; 61 C; 61 G; 65 T; 0 U; 0 Other;
Query Match 9.0%; Score 77.8; DB 4; Length 248;
Best Local Similarity 96.8%; Pred. No. 5.1e-07;
Matches 90; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 169 AAACAGGCTCTTTAAATATACACTTCTCTGTATGCTAGTAAGATGGCTGGA 228
DB 189 AAAATGGCTCTTTAAATATACACTTCTCTGTATGCTAGTAAGATGGCTGGA 131
QY 229 ATAAAAGTGGCTGAATCATCTTCAACTTAGTA 261
DB 130 ATAAAAGTGGCTGAATCATCTTCAACTTAGTA 98
RESULT 13
AAL13109/c
ID AAL13109 standard; cDNA; 534 BP.
XX AC AAL13109;
XX DT 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 5566.
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US000798.
XX PR 14-JAN-2000; 2000US-0176077P.
XX PR 14-MAR-2000; 2000US-0189167P.
XX PR 24-MAR-2000; 2000US-0192099P.
XX PR 29-MAR-2000; 2000US-0193480P.
XX PR 15-MAY-2000; 2000US-0205230P.
XX PR 09-JUN-2000; 2000US-0211315P.
XX PR 25-JUL-2000; 2000US-0220534P.
XX PF (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX DR New peptide useful as a marker for the diagnosis of breast cancer.
XX PT Claim 1; Page 1000-1001; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity
XX SQ Sequence 534 BP; 119 A; 121 C; 119 G; 142 T; 0 U; 33 Other;
Query Match 8.3%; Score 71.6; DB 4; Length 534;
Best Local Similarity 90.4%; Pred. No. 1.1e-05;
Matches 85; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 169 AAACAGGCTCTTTAAATATACACTTCTCTGTATGCTAGTAAGATGGCTGGA 227
DB 265 AAAATGGCTCTTTAAANANACACTTCTCTGTATGCTAGTAAGATGGCTGGA 206
QY 228 ATAAAAGTGGCTGAATCATCTTCAACTTAGTA 261
DB 205 ATAAAAGTGGCTGAATCATCTTCAACTTAGTA 172
RESULT 14
AAS61094
ID AAS61094 standard; DNA; 6062 BP.
XX AC AAS61094;
XX DT 29-JAN-2002 (first entry)
XX DE Human gene regulation-associated gene oligonucleotide #49.
XX KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
XX KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
XX KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
XX KW renal disease; Preeclampsia; cardiac allograft vascular disease;
XX KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
XX KW immunostimulant; cardiac; antiinflammatory; coagulant; antithrombotic;
XX KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

```


PT diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.

PS Claim 1; SEQ ID NO 564; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX

SQ Sequence 6195 BP; 1636 A; 57 C; 1209 G; 3293 T; 0 U; 0 Other;

Query Match 7.7%; Score 66.2; DB 6; Length 6195;

Best Local Similarity 34.2%; Pred. No. 0.00019;

Matches 269; Conservative 0; Mismatches 511; Indels 6; Gaps 3;

Qy 21 TTAAGTTTATGAGATATATCCATGTTGAAATTTGTAGCTGTGGTTTGATGATTTTACAT 80

Db 3391 TTATATTTTATTTTATTTATTTATTTATTTGATGAATATTTAGGTTAATTTATATTTTGT 3450

Qy 81 TATATAGTATACATT-CCATGATAGTTCTCAGTAGATAATCCTCTATTTGTTAACAT 139

Db 3451 TATTTTGAATAATGTTGAAATGGAATATGGGAGTGTAGATTTTATTTTGATATAGTGA 3510

Qy 140 TTGGTTTGTCTTCTCATTTTGACCTATTTTAAACAGGCTCTTTAAATATACACTTCTCTG 199

Db 3511 TAATTTTGGGAGATATTTAGAGTGGATTTGGGTATATGTTAATTTATGTT 3570

Qy 200 TAGTGATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCTGAATCATCTCACTCTAG 259

Db 3571 TAGGTTTGGAGGAATGTTATATTTTAAATGGTTGTTAATTAATTTATATTTTA 3630

Qy 260 TAAGATGTCAACTGTTTCTAAAGTGTATTTTAACTATTTATATATGTCATTTGAA 319

Db 3631 TTAAGGTTTAAAGGTTTATTTTATATTTTATTA--ATACGATTTATTTGTT 3688

Qy 320 CAGCTCTTTCACTTACTAGCAATTTATTCAGCAACACTTGTATTGTCAGACTTTAA 379

Db 3689 TATTTGTTAAAGTATTTTAAATAGATGATTTATTTTAAATTTGATTTT 3748

Qy 380 GTTTCATCTACGNN 439

Db 3749 TTGATAATTAGTATGTCGAGTGTCTTTTATATATTTGCGGTTAGGTTTATGTTT 3808

Qy 440 NNN 499

Db 3809 TTTTAAATTCAGATGTTTGTGTTTATTCAGTGTGTTGAATTTTATATTTTGG 3868

Qy 500 NNN 559

Db 3869 GATATTAGTTTATTTAGATATATGTTTGTAAATATTTTATTTTGGATTTAT 3928

Qy 560 NNN 619

Db 3929 TTTTGTGTTTTTTTGTGTAGAGTTTTTTAGTGTATATTTTGTGTTTATTTT 3988

Qy 620 NNNNNNNNGTATCATATTTATGTTTGTCTCAATTTAAACCTGTTAAATTTTAC 679

Db 3989 TGATTTTGTGTTTGTGTTTGGAGTTTGTGTAGAAATAATTTAGAAAAATGTT 4048

Qy 680 AAATCGACATTATGATATATATCTCCAAATTTTAAATTTTGTCTTTTTCACATT 739

Db 4049 ATAAAGATTTTTTTTATGTTTTTTTGTAGATTTTAAATTTTGGGTTTTTTT---ATG 4105

Qy 740 TTAGTCTTTAGCTCAGCTGGAAATTCATTTCTGTGTGTGTGTAGATAAGTCTTTTTCAT 799

Db 4106 TAAGTTTTTATTTATTTTGTAGTTGAGTTTGTGTAGTGTGGGATAGTGTATTT 4165

Qy 800 GTTTTT 805

Db 4166 TATTTT 4171

Search completed: August 14, 2004, 19:13:42
Job time : 437.526 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 19:13:54 ; Search time 490.203 Seconds
(without alignments)
8638.053 Million cell updates/sec

Title: US-10-082-830-98

Perfect score: 863
Sequence: 1 atgggttggtcattcaact.....gctgagaccatgatagctg 863

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 6451454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	630	73.0	863	15	US-10-082-830-98
2	392	45.4	392	15	Sequence 98, Appl
3	91.2	10.6	1017	15	Sequence 97, Appl
4	89.8	10.4	390	15	Sequence 11216, A
5	89.8	10.4	393	15	Sequence 10053, A
6	89.8	10.4	441	15	Sequence 10053, A
7	89.8	10.4	481	15	Sequence 887, Appl
8	71.6	8.3	1015	15	Sequence 2119, Ap
9	67.6	7.8	6062	13	Sequence 2037, Ap
10	66.2	7.7	6195	15	Sequence 4377, Ap
11	64.4	7.5	1214	13	Sequence 50, Appl
12	63.6	7.4	3673778	15	Sequence 564, Appl
13	62.8	7.3	7128	15	Sequence 102083,
14	62.8	7.3	40324	17	Sequence 1, Appli
					Sequence 1532, Ap
					Sequence 180, Appl

Sequence 71, Appl
Sequence 470, Appl
Sequence 52, Appl
Sequence 40, Appl
Sequence 409, Appl
Sequence 95, Appl
Sequence 107, Appl
Sequence 983, Appl
Sequence 61, Appl
Sequence 52, Appl
Sequence 54, Appl
Sequence 2331, Ap
Sequence 2, Appli
Sequence 836, Appl
Sequence 232, Appl
Sequence 2278, Ap
Sequence 461, Appl
Sequence 2, Appli
Sequence 287, Appl
Sequence 1670, Ap
Sequence 1602, Ap
Sequence 116, Appl
Sequence 325, Appl
Sequence 195935,
Sequence 195935,
Sequence 240, Appl
Sequence 157, Appl
Sequence 1, Appli
Sequence 1280, Ap
Sequence 5, Appli

US-10-311-455-71
US-10-221-714A-470
US-10-172-086-52
US-10-311-507-40
US-10-221-714A-409
US-10-239-676-95
US-10-240-453-107
US-10-311-455-983
US-10-239-676-61
US-10-240-453-63
US-10-239-676-52
US-10-240-453-54
US-10-311-455-2331
US-10-312-841-2
US-10-311-455-836
US-10-221-714A-232
US-10-311-455-2278
US-10-221-714A-461
US-10-204-708-2
US-10-240-453-287
US-10-311-455-1670
US-10-311-455-1602
US-10-257-166-116
US-10-221-613-325
US-10-027-632-195935
US-10-240-453-240
US-10-311-455-157
US-10-668-749A-1
US-10-311-455-1280
US-10-239-676-5

ALIGNMENTS

RESULT 1

US-10-082-830-98
; Sequence 98, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatenIn Ver. 2.1
; SEQ ID NO 98
; LENGTH: 863
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (395)..(627)
; OTHER INFORMATION: a, c, g or t
US-10-082-830-98

Query Match 73.0%; Score 630; DB 15; Length 863;

Best Local Similarity 100.0%; Pred. No. 3.9e-119;

Matches 863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTGGTCATTCACTTTAAGTATATGAGATATCCATGTTGAATTTGTAGCTG 60

Db 1 ATGGGTGGTCATTCACTTTAAGTATATGAGATATCCATGTTGAATTTGTAGCTG 60

QY 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
DB 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
QY 121 TCCTCCATTTGTTTAAACATTTGCGTTCCTCTCATTTTGAACCTATTTTAAACAGGCTCT 180
DB 121 TCCTCCATTTGTTTAAACATTTGCGTTCCTCTCATTTTGAACCTATTTTAAACAGGCTCT 180
QY 181 TTAATATACACTTCTCTGTAGTCTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
DB 181 TTAATATACACTTCTCTGTAGTCTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
QY 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTTAAAGTGTGTTTATTTAACT 300
DB 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTTAAAGTGTGTTTATTTAACT 300
QY 301 ATTATATGTCATTTTGAACAGCTCTTCTCACTTACTAGCAATTTATATCAGCAACACTT 360
DB 301 ATTATATGTCATTTTGAACAGCTCTTCTCACTTACTAGCAATTTATATCAGCAACACTT 360
QY 361 GTTATGTCAGACTTTTAAAGTGTGTTTCACTGCGNNNNNNNNNNNNNNNNNNNNNNNN 420
DB 361 GTTATGTCAGACTTTTAAAGTGTGTTTCACTGCGNNNNNNNNNNNNNNNNNNNNNNNN 420
QY 421 NNN 480
DB 421 NNN 480
QY 481 NNN 540
DB 481 NNN 540
QY 541 NNN 600
DB 541 NNN 600
QY 601 NNN 660
DB 601 NNN 660
QY 661 ACCTGGTTAAATATTTACAAATCGCAATTATGATATTTATCTTCCAAATTTTAAATAA 720
DB 661 ACCTGGTTAAATATTTACAAATCGCAATTATGATATTTATCTTCCAAATTTTAAATAA 720
QY 721 TTTTGTCTTTTTCACATTTTAGTCTTACGCTGGAATTCATTTCTGTGTGGTG 780
DB 721 TTTTGTCTTTTTCACATTTTAGTCTTACGCTGGAATTCATTTCTGTGTGGTG 780
QY 781 TGAGATAAGTCTTTTTCATGTTTTCCTATGAAATAAATTTTCTTCTGTATTGCAG 840
DB 781 TGAGATAAGTCTTTTTCATGTTTTCCTATGAAATAAATTTTCTTCTGTATTGCAG 840
QY 841 GTAGCTGAGCAATGATAGCTG 863
DB 841 GTAGCTGAGCAATGATAGCTG 863

RESULT 2

US-10-082-830-97
; Sequence 97, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: Genes and Proteins
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802

; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-082-830-97

Query Match 45.4%; Score 392; DB 15; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.3e-70; Mismatches 0; Indels 0; Gaps 0;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGTTTGGTCATTCACACTTTAAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTG 60
DB 1 ATGGGTTTGGTCATTCACACTTTAAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTG 60
QY 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
DB 61 TGGTTTGATGATTTTACATTTATAGTATTACATTCATGGATAGTTCTCAGTAGATAA 120
QY 121 TCCTCCTATTGTTTAAACATTTTGCCTTCTCTCATTTTGACCTATTTTAAACAGGCTCT 180
DB 121 TCCTCCTATTGTTTAAACATTTTGCCTTCTCTCATTTTGACCTATTTTAAACAGGCTCT 180
QY 181 TTAATATACACTTCTCTGTAGTCTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
DB 181 TTAATATACACTTCTCTGTAGTCTATGCTAGAAATCGAGTGGCTGGAATAAAGTGGCT 240
QY 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTTAAAGTGTGTTTATTTAACT 300
DB 241 GAATCATCTTCAACTCTAGTAAGATGTCAAACCTGTTTCTTAAAGTGTGTTTATTTAACT 300
QY 301 ATTATATGTCATTTTGAACAGCTCTTCTCACTTACTAGCAATTTATATCAGCAACACTT 360
DB 301 ATTATATGTCATTTTGAACAGCTCTTCTCACTTACTAGCAATTTATATCAGCAACACTT 360
QY 361 GTTATGTCAGACTTTTAAAGTGTGTTTCACTTCACT 392
DB 361 GTTATGTCAGACTTTTAAAGTGTGTTTCACTTCACT 392

RESULT 3

US-10-198-846-11216
; Sequence 11216, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; APPLICANT: Xu, Yongyao
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11216
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1017
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-11216

Query Match 10.6%; Score 91.2; DB 15; Length 1017;
Best Local Similarity 96.9%; Pred. No. 5.9e-09;

```
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 166 TTTAAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCT 225
DB 367 TCTGAAAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCT 426
QY 226 GGAATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 427 GGAATAAAAGTGGCTGAATCATCTTCAACTCTAGTA 462

RESULT 4
US-10-198-846-10053
; Sequence 10053, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10053
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 386, 387, 388, 389, 390
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10053

Query Match 10.4%; Score 89.8; DB 15; Length 390;
Best Local Similarity 97.8%; Pred. No. 7.9e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGA 228
DB 127 AAATGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGA 186
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 187 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 219

RESULT 5
US-10-198-846-887
; Sequence 887, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 887
```

```
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 393
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-887

Query Match 10.4%; Score 89.8; DB 15; Length 393;
Best Local Similarity 97.8%; Pred. No. 8e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGA 228
DB 178 AAATGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGA 237
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 238 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 270

RESULT 6
US-10-198-846-2119/c
; Sequence 2119, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2119
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3, 53
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-2119

Query Match 10.4%; Score 89.8; DB 15; Length 441;
Best Local Similarity 97.8%; Pred. No. 8.3e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 169 AAACAGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGA 228
DB 216 AAATGGCCTCTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGAGTGGCTGGA 157
QY 229 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 261
DB 156 ATAAAGTGGCTGAATCATCTTCAACTCTAGTA 124

RESULT 7
US-10-198-846-2043
; Sequence 2043, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
```

```

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2043
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 2, 3, 4, 5, 13, 14, 16, 17, 21, 27, 37, 48, 74, 93, 101,
; LOCATION: 333, 464, 478
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-2043

Query Match      10.4%; Score 89.8; DB 15; Length 481;
Best Local Similarity 97.8%; Pred. No. 8.6e-09;
Matches 91; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 169 AAACAGGCGCTTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 228
Db 151 AAAATGGCGCTTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGGA 210

QY 229 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 261
Db 211 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 243

RESULT 8
US-10-198-846-4377/c
; Sequence 4377, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4377
; LENGTH: 1015
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: 5, 7, 11, 14, 15, 16, 18, 19, 20, 22, 23, 27, 29, 31, 32,
; LOCATION: 33, 34, 35, 39, 41, 43, 49, 50, 53, 54, 55, 56, 57, 59,
; LOCATION: 61, 62, 63, 65, 67, 69, 73, 75, 81, 82, 83, 84, 85, 86,
; LOCATION: 87, 89, 90, 93, 94, 96, 97, 98, 99, 103, 106, 108, 109,
; LOCATION: 110
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 111, 113, 121, 124, 130, 139, 143, 155, 175, 186, 188, 191,
; LOCATION: 193, 212, 256, 282, 298, 308, 334, 374, 399, 401, 442, 459,
; LOCATION: 462, 472, 495, 497, 498, 502, 512, 515, 516, 569, 585, 602,
; LOCATION: 618, 641, 655, 656, 695, 715, 731, 744, 755, 758, 775
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:

```

```

; NAME/KEY: misc_feature
; LOCATION: 808, 809, 810, 842, 847, 854, 866, 880, 905, 907, 908, 921,
; LOCATION: 924, 925, 926, 933, 934, 937, 939, 940, 951, 957, 958, 963,
; LOCATION: 969, 994, 996, 1013
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-4377

Query Match      8.3%; Score 71.6; DB 15; Length 1015;
Best Local Similarity 90.4%; Pred. No. 6e-05;
Matches 85; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 169 AAACAGGCGCTTTTAAATATACACTTCTCTGTAGTGTATGCTAGAAATGGAGTGGCTGG 227
Db 418 AAAATGGCGCTTTTAAANANACACTTCTCTGTAGTGTATGCTANAAATGGAGTGGCTGG 359

QY 228 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 261
Db 358 ATAAAGTGGCGTGAATCATCTTCAACTCTAGTA 325

RESULT 9
US-10-221-613-50
; Sequence 50, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPNEROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221.613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 50
; LENGTH: 6062
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-50

Query Match      7.8%; Score 67.6; DB 13; Length 6062;
Best Local Similarity 32.8%; Pred. No. 0.00079;
Matches 270; Conservative 0; Mismatches 552; Indels 1; Gaps 1;

QY 20 TTAAAGTTTATGAGATATATCCATGTTGAAATTTTGTAGCTGTGGTTGATGATTTTACA 79
Db 5122 TTTAGTCGCTATATATATTAGTAAATTTTAAAGTAAATTTTGTAGTTAGTTTGTGTTT 5181
QY 80 TTATATAGTATATACATCCATGGATAGTCTCTCAGTAGATAATCCTCTATGTTTAAACAT 139
Db 5182 TTGATATAAATTTATAGTTTATGTTTAAATATTTGTTTGGATGTTTAAAGAT 5241
QY 140 TTGCGTTCTCTCTCATTTTGGACCTTATTTAAACAGGCGCTTTTAAATATACACTTCTG 199
Db 5242 ATTTTAAAAATTTTAAATTTAAATTTTATTTTATTAGTTTATTTATGTTTAAATTTTAA 5301
QY 200 TAGTGTATGCTAGAAATGGAGTGGCTGGATAAAGTGGCTGAATCATCTTCAACTCTAG 259
Db 5302 AAGTGGGTATTATTAAT-TCGTTTTTGTAAATTTAGAAATATGGGTATTTATTATGTCGT 5360

```


LENGTH: 1214
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1214)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_WRT3847_63198C.1
US-10-424-599-102083

Query Match 7.5%; Score 64.4; DB 13; Length 1214;
Best Local Similarity 29.5%; Pred. No. 0.0019;
Matches 245; Conservative 0; Mismatches 566; Indels 0; Gaps 0;

QY 6 TTTCGTCATTCACCTTTAAGTTTATGAGATATATCCATGTTGAATTTGTAGCTGGTT 65
Db |||||
QY 66 TGATGATTTTACATTTATAGTATTACATTCATGATAGTTCTCAGTAGATAAATCCCTC 125
Db |||||
QY 959 TTTTCTTTTCTTTTATTTTATTTTGTATTCCTTTTATTTTATTTTATTTTATTTT 900
QY 126 CTATGTTTACATTCGGTTCCTCTCATTTTGACCTATTTTAAACAGCCCTCTTAA 185
Db |||||
QY 899 TTTATTTTATTTGTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 840
QY 186 TATACACTCTCTGCTAGTATGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCTGAATC 245
Db |||||
QY 839 TTTTCTTTTCTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 780
QY 246 ATCTTCAACTCTAGTAAGATGTCAACTGTTTCTTAAAGTGTATTTTAACTATTAT 305
Db |||||
QY 779 TTTTCTTTTCTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 720
QY 306 ATGTCATTTTGAACACTCTTTCACCTTACTAGCAATTTATATCAGCACACTTGTAT 365
Db |||||
QY 719 TAGTCTATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 660
QY 366 TGTGAGACTTTTAAAGTTTCACTTCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 425
Db |||||
QY 659 TGTGTAATTTCAATTTAATTTGTTACATCTATAAGAGAGCTTTATGTAAGTTTGCAAGC 600
QY 426 NNN 485
QY 599 ACGTGAATGTAAGCTCTGAAATCTGTCCTTGTATGTTTCGAGTTATCTCTCATTTCTCGAT 540
Db |||||
QY 486 NNN 545
QY 539 TGTGCTCAGAGAAAGTTAGAAAGCTTTACAGAAAGATGAGCCCAAGCAATCTTT 480
QY 546 NNN 605
QY 479 GACAAATAATAGATTAATTTGTTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 420
QY 606 NNN 665
QY 419 GTTTATGTTTATTTGATGTTTATTTTAAATTTTGTATTTATTTTATTTTATTTTATTT 360
QY 666 GTTAAATATTTCACAAATGACATTTATGATATATCTTCCAAATTTTAAATTTTATTTT 725
Db |||||
QY 359 TTTAAATATTNTTTTAAATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 300
QY 726 TCTTTTTCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 785
Db |||||
QY 299 ATTTTCTTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 240
QY 786 TAAGTCTTTTTCATGTTTCTTCCCTATGAAATAATTTATTTCTTCTGTTAT 836
Db |||||
QY 239 TTTTCTTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 189

RESULT 12

US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 7.4%; Score 63.6; DB 15; Length 3673778;
Best Local Similarity 29.3%; Pred. No. 0.062;
Matches 171; Conservative 0; Mismatches 412; Indels 0; Gaps 0;

QY 276 TTTCTAAAGTGTTTTATTTTAACTATTATATGTCAAATTTTGAACAGCTCTTTCACCTTAC 335
Db |||||
QY 1241299 TGTGTAATAGTTTATTTTGTAGTATTTTGTATTTTATTTTATTTTATTTTAA 1241358
QY 336 TAGCAATTTATATCAGCAACACTTGTATTGTCAGACTTTTAAAGTTTTCATTCACCTGGN 395
Db |||||
QY 1241359 TCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1241418
QY 396 NNN 455
Db |||||
QY 1241419 TTTTCTTATGTTTGGAGAAAAAAGTTTAGAAGAAATATAAATATTTTATTTAT 1241478
QY 456 NNN 515
QY 1241479 ATGTTTGTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1241538
QY 516 NNN 575
QY 1241539 TGTGTTGTTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1241598
QY 576 NNN 635
QY 1241599 AGTAGTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATTTAT 1241658
QY 636 ATTTATGTTTGTCTCAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 695
QY 1241659 TTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1241718
QY 696 ATATTATCTTCCAAAAATTTTAAATTTTGTCTTTTTCATTTTATTTAGTCTTTAGTCTCAG 755
QY 1241719 TTTTATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1241778
QY 756 CTGAATTCATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 815
QY 1241779 TAGGTATTTATTTGTTGTTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1241838
QY 816 TAAATTTTCTCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 858
QY 1241839 TTTTATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1241891

RESULT 13

US-10-311-455-1532
; Sequence 1532, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: FIEPENBROCK, Christian

[illegible]

```
Db 38536 AGTTTTCGTAGATTAAAGAGTTTTTTTTATGAAGTTTTTTTTTGTGTTTTTTTTT 38595
Qy 661 ACCGTGTTAAATATTTCCACAAATCGACATTATGATATATATCTCCAAAATTTTATATA 720
Db 38596 TATTTTTATATGTTTGTGTTTTTTTAAATGAAAATCGAATATGTAATATATGTAAGA 38655
Qy 721 TTTTGTCTTTTTTACAAATTTAGTCTTTAGCTCAGCTCGAATTCATTTCTGTGTGGTG 780
Db 38656 TTTTGCAGTTTATCGGGTTTTTTTTTTTGTAGCTATTAATTTATTTTATGTTTGTGTTTG 38715
Qy 781 TGAGATAAGTCTTTTCAAGTTTTTCCCTATGAATAAATTAATTTCCCTCTGTAT 835
Db 38716 TTTGATAGTTAAATTTGTTCCGAATTAATAATGTTAAATATGTTTTTATAGGGTAT 38770

RESULT 15
US-10-311-455-71
; Sequence 71, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 71
; LENGTH: 9964
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-71

Query Match 7.0%; Score 60.6; DB 15; Length 9964;
Best Local Similarity 33.2%; Pred. No. 0.026;
Matches 275; Conservative 0; Mismatches 547; Indels 6; Gaps 2;

Qy 2 TGGGTTTGGTCATTCAACTTTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGT 61
Db 8315 TGTGTTTTTAAATATATAAGTTGACGGTGTAGTAGATTTTATTTAGTAAATTTTTTTTATA 8374
Qy 62 GGTGTTGATGATTTTATACATATATAGTATATACATTCATCGATGATTCAGTAGATAAT 121
Db 8375 GTGTAGTGTGTTTATGTTTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 8434
Qy 122 CCTCCTATGTTTAAACATTTTGGTGTCTCTCTCATTTTGAACATTTTAAACAGGCTCTT 181
Db 8435 TTTCGTTTATTTTGTGTTTGTGTTTTTTTAAATTTGTTTTTTTTTTTATTTGTTTTTTT 8494
Qy 182 TAAATATACATCTCTCTGTAGTGTAGCTAGAAATGGAGTGGCTGGAATAAAAGTGGCTG 241
Db 8495 TTTTTCGTTTTTTTGTAGTGTGTTTTTTTGGTGTGTTTTTGTAGTTTTTATTTTTTAAAT 8554
Qy 242 AATCATCTTCAACCTCTAGTAGATGTCAAACGTTTTTCTAAAAGTGTGTTTTTATTTAACTA 301
Db 8555 TTATTTTTTTTGTGTTTATTTTGTGTTTTTTTATTTTGTGTTTTTGTATTTGTTTTTTTAA 8614
Qy 302 TTATATGTCATTTTGTGAACAGCTCTTTCACTTACTAGCAATTTATATCAGCAACACTTG 361
Db 8615 TTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGT 8672
Qy 362 TTATGTCAGACTTTTAAAGTTTTTCACTCACTGGNNNNNNNNNNNNNNNNNNNNNNNNNN 421
```

```
Db 8673 TTTTGTGTTTTTAAATTTATTTAGTTTTTTTGTAAATTTTGGTAGTATTTTATGTTTGTGTT 8732
Qy 422 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 481
Db 8733 TGAATTGATTTTATTTATTTGTTGGATTTTATTTATTTTGTGTTTTTATTTATGAAATTTT 8792
Qy 482 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 541
Db 8793 TGTATAAATTTTGTAGATTTATTTGATTTTTTTTATGTTTTTTTCGTGTTTTTGTGTTAT 8852
Qy 542 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 601
Db 8853 TTGTTTTGTTTTTATTTTGTGTTGTTATTTATTTATTTATTTTATTTTATTTTATTTTAT 8912
Qy 602 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 661
Db 8913 TTTTTCGATTTAAATTTTGTGTTTTTATTTATTTATTTGTTTTTTTATTTTATTTTATTTT 8972
Qy 662 CCTGTTAAATTTTCAAAATCGACATATGATATATATCTTCCAAAATTTTAAATAT 721
Db 8973 GATTTTTTTTTTTTTTAAATTTTGTGTTTTTAAATTTATTTATTTATTTATTTATTTAT 9028
Qy 722 TTTGTCTTTTTTCACATTTTGTAGCTTTAGCTCAGCTGCAATTCATTTCTGTGTGGTGT 781
Db 9029 TTTGTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTAAATTTTGTGTTTATTTATTTATTTAT 9088
Qy 782 GAGATAAGTCTTTTTCATGTTTTTCCCTATGAAATAAATTTATTTCCCTT 829
Db 9089 TTTTTCGATTTTTTTTTTTTTTTTTTTTTTAAATTTTGTGTTTATTTATTTATTTATTTAT 9136

Search completed: August 15, 2004, 01:59:51
Job time : 495.203 secs
```


Db 2579 GTAAAGGTTTATAATTTTATTTTATTTTATATGTTAAATTTAGTGGTATATTTT 2638
 QY 282 AAAGTGTTTTAACTATATATGCAATTTTGAACAGCTCTTTCACCTACTAGCAA 341
 Db 2639 ATAGTTATTTATTAATTTTGAATTTTATTTTATAGAGTTTAAATGTTATTTGTT 2698
 QY 342 TTATATACAGCAACTGTTATGTCAGACTTTTAAAGTTTCAITTCACCTGNNNNNN 401
 Db 2699 TAGGAGATTTTGAATTTTATTTTATTTTATGTTTGGGGTTTATTTTATGAGTTT 2758
 QY 402 NNN 461
 Db 2759 TTATTTAGTATTTGTTAAATTTATTTTATTTTTCGTTATAGAGTTATGGAATTTT 2818
 QY 462 NNN 521
 Db 2819 TTTTATGCTATATATATTTGTTAGTGTGTTGAAGGTAATTAATATTAGTGTGTTTATA 2878
 QY 522 NNN 581
 Db 2879 AGTTATATTTTATTTTATTTTAAATTTTAAATAGATATTGATGAGGGTTTAGT 2938
 QY 582 NNN 641
 Db 2939 TGATTTTATGCTGAGTTTGGGGGTTAAGGATTTATGTTTAAATTTTATTTTATTA 2998
 QY 642 GTTTTGTCTATTTAAACCACTGGTTAAATATTTCAAAATCGACATTAATGATATTA 701
 Db 2999 GTATTTGGTAAAGTTGATTTTAAATATTTGTTGAATTAATGATTTGATGAATAA 3058
 QY 702 TCTTCCAAATTTTAAATTTTGTCTTTTTCACATTTTATGCTTTAGCTCAGCTGGA 761
 Db 3059 GACGTTATTAATGTAAGTAAATTTTATTTTATTTTGTGTTTATTTTACGTTAT 3118
 QY 762 TCAATTTCTGTGTGTGTGAGATAAGTCTTT 794
 Db 3119 TATATTTAGGATAGAGGTAATGTTGTAT 3151

RESULT 2

US-10-204-708-39
 ; Sequence 39, Application US/10204708
 ; Patent No. 667731
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
 ; FILE REFERENCE: 5013.1012
 ; CURRENT APPLICATION NUMBER: US/10/204,708
 ; CURRENT FILING DATE: 2003-05-06
 ; PRIOR FILING DATE: 2001-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019058.8
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019173.8
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 39
 ; LENGTH: 19513
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-204-708-39

Query Match

Best, Local Similarity 5.9%; Score 50.8; DB 4; Length 19513;
 Pred. No. 0.0071;

Matches 241; Conservative 0; Mismatches 550; Indels 0; Gaps 0;
 QY 35 TATATCCATGTTGAATTTTGTAGCTGTGCTTTGTAGTATTTTACATATATAGTATTACA 94
 Db 13792 TTTTAAATTTTATGTTTAAATTTAGTGGTATTTATATGCGGTGTTATGTTAGGTT 13851
 QY 95 TTCCATGGATAGTCTCAGTAGATAATCCTCTATTTGTTTAAACATTTTCGTTGCTTCTCA 154
 Db 13852 TTTAGTGTATAGTTAGTTGTATAGTGTGTTTATTTTATTTAGAAAATTTTATATTTT 13911
 QY 155 TTTTACACTATTTTAAACAGGCTCTTTAAATATACACTTCTCTGTAGTGTATGCTGAA 214
 Db 13912 TTTAGTGTATAGTTAGTTGTATAGTGTGTTTATTTTATTTAGCGTTAGATTAATTT 13971
 QY 215 ATGGAGTGGCTGGAATAAAAGTGGCTGAATCATCTTCAACTCTAGTAAGATGTCAAACTG 274
 Db 13972 TAGTGTATTTTATAATTTTGTTTAAATGCTATGTAATTTTATGAGGTAATAATAG 14031
 QY 275 TTTTCTAAAGTGTTTTATTTTAACTATATATGTCATATTTTGAACAGCTCTTTCACTTA 334
 Db 14032 TTATTTTAAAGTAATAATGTTGGGCTGTGTTTATATTTGTAATTTTAGTATTTTG 14091
 QY 335 CTAGCAATTTTATATCAGCAACTTGTATTTGTCAGACTTTTAAAGTTTTCATTCACCTG 394
 Db 14092 GGGAGTTAAGGTAGGTAGATTAATTTGAGTTTAGGGTTTTAGTATTTTAGTAATAT 14151
 QY 395 NNN 454
 Db 14152 GGTAAATTCGTTTTTATTAATAAAATAATTTGGTTAGGTATGTTGCTACGTATTGTTT 14211
 QY 455 NNN 514
 Db 14212 ATAATTTTAGAGGTTAAGGTGGGAGGATTAATTTGAGTTTCGGGAGGTTAAGGTTGTAG 14271
 QY 515 NNN 574
 Db 14272 TGAGTGTAGATTAATGTTATTTAGTTTGGGTAATAGAGTGAGATTTTGTTTAAT 14331
 QY 575 NNN 634
 Db 14332 AAAGAGAGAAGTAATAATAAGTAAGAAATTTAGTTGGTACGTTTATAGTGTGTATAA 14391
 QY 635 TATTTATGTTTGTCTCATTTTAAAAAACCTGGTTAAATATTTTCAAAATCGACATTAAGA 694
 Db 14392 TTATATTTTGTAGTTTAAATAATTTTATTTTATAGAAAATTTTATATTTTCG 14451
 QY 695 TATATATCTTCCAAATTTTAAATAATTTTGTCTTTTTCACATTTTGTCTTTAGTCTTTAGCTCA 754
 Db 14452 TAAGTAGTATTTTATTTTATTTTATTTTGTGTTATTTTGTGTTATAGTATGTTTAA 14511
 QY 755 GCTGGAATTCATTTCTGTGTGTGTGAGATAAGTCTTTTTCATGTTTTCCTCATGAA 814
 Db 14512 GATTTATCTGTGTGTGTGTGTTGTTAGGAGTTGTTTTTTTAAATGTCAGTGTGTT 14571
 QY 815 ATAAATTTATTT 825
 Db 14572 ATGAATTTTATTT 14582

RESULT 3

US-10-204-708-18
 ; Sequence 18, Application US/10204708
 ; Patent No. 667731
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
 ; FILE REFERENCE: 5013.1012
 ; CURRENT APPLICATION NUMBER: US/10/204,708
 ; CURRENT FILING DATE: 2003-05-06
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03971

```

; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 18
; LENGTH: 5535
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; US-10-204-708-18

Query Match          5.7%; Score 49.2; DB 4; Length 5535;
Best Local Similarity 27.4%; Pred. No. 0.013;
Matches 153; Conservative 0; Mismatches 406; Indels 0; Gaps 0;

Qy 276 TTTCCTAAAGTGTTTATTTTAACTATATATGTCAAATTTTGAACACCTCTTTCACCTTAC 335
Db 1262 TTTATATTTTGTGTTATTTTAAATGGGTGTGATGGTTTTTTTATTATTATTAATTTT 1321
Qy 336 TAGCAATTTATATCAGCAACACTGTGTTATGTGCAGACTTTTAAAGTTTTCATTCACTG 395
Db 1322 TAGGGGTATTTATGGATTGTGATTTAAATTTTAGTTATTATATGTAATAATTTT 1381
Qy 396 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 455
Db 1382 TTTAGTTTATTTTGGGTTTAAATTTTAGTTATTTTGTGTATAGAAAGTTATTGTGTTT 1441
Qy 456 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 515
Db 1442 ATCGGTAATTATTTATTAATTTGTTGTTTCGTTTGTGATTTTGTGTTGTTGTTAAGATT 1501
Qy 516 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 575
Db 1502 TTGTTATATTTTACGGTTATATAATCATATAATTTTATTTTATTTTATTTTTCGA 1561
Qy 576 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 635
Db 1562 GACGGAGTTTCGTTTTTATTTATTTAGGTTGGAGTGTAATGCGCGGATTTTAGTTTATGT 1621
Qy 636 ATTTATGTTTTCGTCATTTAAAAAACCCTGGTTAAATATTTTCACAAAATCGACATATTGAT 695
Db 1622 ATTTTATTTTTCGGGTTTAGCGATTTTGTGTTTTTTAGTAGTTGGGATATAGGTA 1681
Qy 696 ATATATCTTCCAAAATTTTAAATATTTTGTCTTTTTTTCACATTTTAGTCTTTAGCTCAG 755
Db 1682 TGTGTTATTATGTTCCGGTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTAGTAG 1741
Qy 756 CTGGAAATTCATTTCTGTCGTGGTGATAGATAAGTCTTTTTTCATGTTTTTCCCATGAAA 815
Db 1742 ACGGGGTTTATATGTTGGTTAGTTCGTTGTTTTAAATTTTCAITTTTAGTTGATTTTGAG 1801
Qy 816 TAAATTTATTCCTCTGTA 834
Db 1802 GTTATTTTGTGATTTTTTA 1820

```

```

RESULT 4
US-10-204-708-89
; Sequence 89, Application us/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: FIEBENEROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: By Assessing DNA Methylation

```

```
; FILE REFERENCE: 5013_1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 89
; LENGTH: 5844
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-89

Query Match          5.6%; Score 48.6; DB 4; Length 5844;
Best Local Similarity 27.8%; Pred. No. 0.018;
Matches 159; Conservative 0; Mismatches 417; Indels 0; Gaps 0;

Qy      275 TTTTCTAAAGTGTTTTATTTAACAATTATAATGTCACACTTTTGAACAGCGCTTTTCACTTA 334
Db      4031 TAATTGATATAGTATAAAGTTAGTTATTATTTTAGTTTGAACATATATTTATGATTCG 4090

Qy      335 CTAGCAATTTATTATCAGCAACACTGTATTGTCAGACCTTTTAAGTTTTCATTCACCTGG 394
Db      4091 GAAGAAGAGTTTTTATGATTAAAGTAGTAGTTTTTATTTTATTTTATATTTATTTA 4150

Qy      395 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 454
Db      4151 TTATTATTATTTTTTTTAGGAGTTTCGTTTTCGTTAGTTCGAGTGTAGTGTATAAATT 4210

Qy      455 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 514
Db      4211 TCGGTTTATGTAATTTTCGTTTTCCGGFTTAAAGTTATTTTTTGTGTTTAGTTTTTGA 4270

Qy      515 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 574
Db      4271 GTAGTCGGATTATAGCGGTAGTTCGTTACGTTCCGGTTAATTTTTTTTTTTTTTTGTATT 4330

Qy      575 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTATCA 634
Db      4331 TTAGTAGAGATGGGTTTTTATTATGTTGTTAGGTGGTTTCGAAATTTTGAGTTTAGGT 4390

Qy      635 TATTATGTTTTCGTCAATTTAAAAAACCCGGTTAAATAATTTACAAAATCGACATATGA 694
Db      4391 AGTTATTATTTTAGTTTTTAAAGTGTAGGATTATAGGTATGCGTTATCGTATTTCGG 4450

Qy      695 TATATATCTTCCAAAATTTTAAATTTTGTCTTTTTCACATTTTAGTCTTTAGCTCA 754
Db      4451 TAGTTTTTTATTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTGTGT 4510

Qy      755 GCTGGAATTCATTTCTGTGTGTGTGCATGAAGTCTTTTCAIGTTTTTCCCTATGAA 814
Db      4511 TTAGTTGGAGGTGAGTCACAGAGATTTAGTTATTATGTAATTTTGTTTTTTAGGTTAA 4570

Qy      815 ATAAATATTTCCTCTGTATTCAGGTAGCTGAGA 850
Db      4571 GCGATTTTTTTGTTTATGTTTTTTTGTAGTTTTTTGAGTAGTTCGGA 4606
```

RESULT 5
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir

APPLICANT: Jeffery Gulcher
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
FILE REFERENCE: 2345-2001-000
CURRENT APPLICATION NUMBER: US/09/426,290
CURRENT FILING DATE: 1999-10-25
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 168575
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match 5.6%; Score 48.2; DB 4; Length 168575;
Best Local Similarity 59.2%; Pred. No. 0.044; Indels 1; Gaps 1;
Matches 100; Conservative 0; Mismatches 68;
Qy 11 TCATTCACTTTAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGTGTTTATG 70
Db 13525 TCATTGAGTTTAAATTTTCCAGTTTCTTCCAGTTTATGTTGCTATCCGTTATTA 13466
Qy 71 ATTTTACATATATAGTATATACATCCATGAGTAGTTCTCAGTAGATA-ATCCTCTAT 129
Db 13465 TTTTGTGGCTGATAGTATTTATATATAGATATGCCACATTTATCTATTCAGCAT 13406
Qy 130 TGTTAACATTTGGTTGCTTCTCATTTGACCTATTTTAAACAGGCCT 178
Db 13405 TGATGAATTTGGATTGTTTCCATTTTGGCAATATATATAAATACT 13357

RESULT 6
US-08-559-896B-1/c
Sequence 1, Application US/08559896B
Patent No. 6310046
GENERAL INFORMATION:
APPLICANT: Patrick E. Duffy
APPLICANT: Christian F. Ockenhouse
TITLE OF INVENTION: SEQUESTIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John Moran
STREET: USA MRC - MCMR-JA
CITY: FORT DETRICK, FREDERICK
STATE: MARYLAND
COUNTRY: USA
ZIP: 21702-5012
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.5
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,896B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Moran, John
REGISTRATION NUMBER: 26,313
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 619-2065
TELEFAX: (301) 619-7714
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1956 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: Linear
US-08-559-896B-1
Query Match 5.6%; Score 48; DB 4; Length 1956;
Best Local Similarity 26.5%; Pred. No. 0.02; Indels 0; Gaps 0;
Matches 138; Conservative 0; Mismatches 383;
Qy 305 TATGTCAATTTGAAACAGCTCTTTCACTTACTAGCAATTTATATACGACACTTGTTA 364
Db 1360 TGTCTAAATTTGTAACAATAATCAACTTCTACATTTTCATTATAATTTATTTGT 1301
Qy 365 TTGTCAGACATTTTAAGTTTTCATTCAGTGGNNNNNNNNNNNNNNNNNNNNNNNN 424
Db 1300 GCTTATTATTTAAATTAATTCACATCTCTATTAATTAATTAATTAATTAATCAT 1241
Qy 425 NNN 484
Db 1240 TATTGTCGTGATTATAAATACATTTTCTTTCTTTATTTTGGTTAGATTCAT 1181
Qy 485 NNN 544
Db 1180 TTTTAATTTCTATGATATAATATATCTTTTTCATTTCTGCTTAAATTTCTCATGGGA 1121
Qy 545 NNN 604
Db 1120 TTTTCATACCTTCTCTATCCATTTCTATCTAATCTCTTTTATAAGCTCATCGATT 1061
Qy 605 NNN 664
Db 1060 CTTCTAGTTCTAAATTTTGTATTTTCTTTTGTATATCTGTAAATTTGCTCTCTTA 1001
Qy 665 GGTAAATATTTCAAAATCGACATGATATATATCTTCCAAAATTTTATAAATTTT 724
Db 1000 CATGTGAATTTCACTATCCATTTGTTGTTAAATTTCTTCTATATTTTATCTATT 941
Qy 725 GTCTTTTTCACATTTTAGTCTTTTAGCTAGCTGGAATTCATTTCTGTGTGTGTGAG 784
Db 940 GATCTTTATCCATTTTATTTTCTTTTCTTTTCAATTTTGTGTAATTTTCTCTATGA 881
Qy 785 ATAAGTCTTTTTCATGTTTTCCTTATGAATAAATTTT 825
Db 880 TTTTCTTTTCTCTCTGTTCTATTTTCTAATTTCTCT 840

RESULT 7
US-09-351-794A-1/c
Sequence 1, Application US/09351794A
Patent No. 664815
GENERAL INFORMATION:
APPLICANT: DUFFY, PATRICK E.
APPLICANT: OCKENHOUSE, CHRISTIAN F.
TITLE OF INVENTION: SEQUESTIN
FILE REFERENCE: 38644-175519
CURRENT APPLICATION NUMBER: US/09/351,794A
CURRENT FILING DATE: 1999-07-13
PRIOR APPLICATION NUMBER: 09/559,896
PRIOR FILING DATE: 1995-11-17
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1956

```

; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-351-794A-1
Query Match      5.6%; Score 48; DB 4; Length 1956;
Best Local Similarity 26.5%; Pred. No. 0.02;
Matches 138; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

QY 305 TATGTCAAATTTCAACAGCTCTTTCACCTACTAGCAATTTATATACAGCAACACTGTGTA 364
Db 1360 TGCTAAATTCGTCAACAACTAATCAACTCTACATTTTCATATATATTTATTGTTGTT 1301
QY 365 TTCTCAGACTTTTAAAGTTTTCATTCCTACGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 424
Db 1300 GCTTATTTAAATTTATTCACATCTCTATTTATTTATTTATTTATTTATTTATTTAT 1241
QY 425 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 484
Db 1240 TATTTGTCGTGTAATTATAAATACATTTTTCATTTTCTTTCTTTTATTTTGGTTAGATTCAT 1181
QY 485 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 544
Db 1180 TTTTAAATTTCTATGTATATATATATCTTTTTCATTTCTGCTTAAATTTCTCATGGGA 1121
QY 545 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 604
Db 1120 TTTTCATACCTTGTCTCTCTATCCATTTCTATCTAATTTCTTTTATAAAGTCTATCGATTT 1061
QY 605 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 664
Db 1060 CTCTAGTCTTAAATTTTGTATTTTCTTTTCTTTTGTATATCTGTAATGCTCTCTTA 1001
QY 665 GGTTAAATATTTCAAAATCGACATTTATATATATCTTCCAAAATTTTAAATAATTTT 724
Db 1000 CATGTTGAATTTTCATCACTATCCATTTTGTATTTCTTTCTCTCATATTTTATCTATTT 941
QY 725 GTCTTTTTCACATTTTAGTCTTTAGCTCAGCTGGAATTCATTTCTGCTGCTGCTGAG 784
Db 940 GATCTTTTATCCATTTTATTTTATTTTCTCTTTTCTCAATTTTGTGTAATTTTCTCTAGAA 881
QY 785 ATAAGTCTTTTTCATGTTTTTCCCTATGAATAAATTTT 825
Db 880 TTTTTCCTTTTCTCTCTGTCATTTTCTTAAATCTTCT 840

RESULT 8
US-09-539-333D-1/c
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.04TAUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452

; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.Pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc_feature
```

```

; LOCATION: 213818..215818
; OTHER INFORMATION: 3' regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231870..231879
; OTHER INFORMATION: exon O1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 234174..234321
; OTHER INFORMATION: exon O complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 237406..237428
; OTHER INFORMATION: exon Nbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239807
; OTHER INFORMATION: exon N2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239853
; OTHER INFORMATION: exon N complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240569
; OTHER INFORMATION: exon M1117 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240596
; OTHER INFORMATION: exon M1090 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240617
; OTHER INFORMATION: exon M1069 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240644

```

```

; OTHER INFORMATION: exon MS2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240824
; OTHER INFORMATION: exon M862 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240994
; OTHER INFORMATION: exon M692 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..241685
; OTHER INFORMATION: exon M1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240800..240993
; OTHER INFORMATION: exon MS1 complement g34872 gene
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 241686..243685
; OTHER INFORMATION: 5' regulatory region g34872 gene
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 290652..292652
; OTHER INFORMATION: 3' regulatory region g34665 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 292653..292841

```

```

Query Match      5.4%; Score 46.6; DB 4; Length 319608;
Best Local Similarity 65.8%; Pred. No. 0.12;
Matches: 98; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

```

```

QY 27 TTATGAGATATATCCATGTTGAATTTTGTAGCTGTGTTTGATCATTTTAC-ATTATAT 85
    |||||
Db 15622 TTTTGAGTTTCATCCCAIGTTGCTGGGTGTAGCAATACTACTCATTTTAACTGGTCT 15563

QY 86 AGTATTACATTCATGGATGTTCTCAGTAGAT-AATCCTCCTATTGTTTAAACATTTCG 144
    |||||
Db 15562 AGAATTCATTCATGGATATACCAAAATCCATCCATCTTACTGTGTTGATGACATTGG 15503

QY 145 TTGCTTCTCATTTTGACCTATTTTAAACA 173
    |||||
Db 15502 CTTTTCCTTTGGAGCTATTCAAATA 15474

```

```

RESULT 9
US-09-679-409-1/c
; Sequence 1, Application US/09679409
; Patent No. 6555316
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Essieux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US15.CIP
; CURRENT APPLICATION NUMBER: US/09/679,409
; CURRENT FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 09/416,384
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 60/168,088
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: 199122..201122
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 201123..201234
; OTHER INFORMATION: exon S
; NAME/KEY: exon
; LOCATION: 201123..201560
; OTHER INFORMATION: exon S2
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U
; NAME/KEY: exon
; LOCATION: 216836..216994
; OTHER INFORMATION: exon V
; NAME/KEY: exon
; LOCATION: 216836..217077
; OTHER INFORMATION: exon V2
; NAME/KEY: exon
; LOCATION: 217671..217764
; OTHER INFORMATION: exon V1
; NAME/KEY: exon
; LOCATION: 227655..227736
; OTHER INFORMATION: exon V4
; NAME/KEY: exon
; LOCATION: 238715..238919
; OTHER INFORMATION: exon V3
; NAME/KEY: exon
; LOCATION: 240440..240673
; OTHER INFORMATION: exon W
; NAME/KEY: exon
; LOCATION: 240440..241153
; OTHER INFORMATION: exon W2
; NAME/KEY: exon
; LOCATION: 241072..241291
; OTHER INFORMATION: exon X
; NAME/KEY: exon
; LOCATION: 244353..244561
; OTHER INFORMATION: exon Y
; NAME/KEY: exon
; LOCATION: 246273..247802
; OTHER INFORMATION: exon Z
; NAME/KEY: misc feature
; LOCATION: 247803..249803
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 8316
; OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 21672
; OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 65485
; OTHER INFORMATION: 8-128-33 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 95396
; OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 107281
; OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 160640
; OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 160876
; OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 168974
; OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 170810
; OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 173358
; OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 189357
; OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 197163
; OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 200778
; OTHER INFORMATION: 8-303-235 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 202851
; OTHER INFORMATION: 8-300-221 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 202679
; OTHER INFORMATION: 8-300-193 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 203378
; OTHER INFORMATION: 8-299-128 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 204138
; OTHER INFORMATION: 8-296-213 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 204605
; OTHER INFORMATION: 8-252-190 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 204934
; OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 205206
; OTHER INFORMATION: 8-295-248 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 205329
; OTHER INFORMATION: 8-295-125 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 206064
; OTHER INFORMATION: 8-293-130 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 206545
; OTHER INFORMATION: 8-292-198 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 207313
; OTHER INFORMATION: 8-251-322 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 208285
; OTHER INFORMATION: 8-289-322 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 208960
; OTHER INFORMATION: 8-287-249 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 209123
; OTHER INFORMATION: 8-287-86 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 209631
; OTHER INFORMATION: 8-285-319 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 210361
; OTHER INFORMATION: 8-283-278 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210463
; OTHER INFORMATION: 8-283-176 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 210486
; OTHER INFORMATION: 8-283-153 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210583
; OTHER INFORMATION: 8-283-56 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 210879

```



```

; OTHER INFORMATION: 8-282-345 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210964
; OTHER INFORMATION: 8-282-260 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 210979
; OTHER INFORMATION: 8-282-245 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 211050
; OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
; NAME/KEY: allele
; LOCATION: 211132
; OTHER INFORMATION: 8-282-92 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 211247
; OTHER INFORMATION: 8-281-367 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 211315
; OTHER INFORMATION: 8-281-299 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 211366
; OTHER INFORMATION: 8-281-248 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 212520
; OTHER INFORMATION: 8-279-197 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 212821
; OTHER INFORMATION: 8-278-289 : polymorphic base C or T
; NAME/KEY: allele

```

```

Query Match 5.4%; Score 46.6; DB 4; Length 319608;
Best Local Similarity 65.8%; Pred. No. 0.12;
Matches 98; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

QY 27 TTAAGATATATCCATGTTGAATTTTGTAGCTGTGGTTGATGATTTTAC-ATTATAT 85
Db 15622 TTTTGTAGTTTCATCCATGTTGCTGGGTAGCAATAATCTACTCATTTTAACTGTGCT 15563

QY 86 AGTATTACATCCATGATAGTTCTCAGTAGAT-AATCTCTCTATTGTTTACATTTGG 144
Db 15562 AGAATCCATCTATGATATACCAAAATCCATCCATTTCTACTGTGTGATGACATTTGG 15503

QY 145 TTGCTCTCTATTTTGACCAATTTTAAACA 173
Db 15502 CTTTTCCATTTGGAGCTATTTCCAAATA 15474

```

```

RESULT 10
US-08-951-871-1/c
; Sequence 1, Application US/08951871
; Patent No. 5866398
; GENERAL INFORMATION:
; APPLICANT: XU, SHUANG-YONG
; APPLICANT: XIAO, JIAN-PING
; TITLE OF INVENTION: METHOD FOR CLONING
; TITLE OF INVENTION: AND PRODUCING THE BclI RESTRICTION ENDONUCLEASE IN E.
; NUMBER OF INVENTION: COLI
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,871
; FILING DATE:

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Gregory D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 978-927-5054
; TELEFAX: 978-927-1705
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2739 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..2736
; OTHER INFORMATION:
; US-08-951-871-1
;
Query Match 5.4%; Score 46.2; DB 2; Length 2739;
Best Local Similarity 47.2%; Pred. No. 0.054;
Matches 141; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 91 TACATTCATGATAGTTCCTCAGTAGATAATCTCTCTATTGTTTAACTTTGGTTCCTT 150
Db 917 TAAATTTGTTTATGGAGAAAGTACATAATCAAGAAATTTATTTTCATCTTGAATCTA 858

QY 151 CTCATTTTGACCTATTTTAAACAGGCTCTTTAAATATACACTTCTCTGTAGTGTATGCT 210
Db 857 TTAATAAAATCTCTATCTATATGGAATAAAACAAAGTCATCTCTCTGTAGTAAGAGT 798

QY 211 AGAATGGAGTGGTGAATAAAGTGGCTGAATCATCTTCAACTCTAGTAAGATGTCAA 270
Db 797 AGAATATCATTTCCCGGTCAAAATAAATAAGGATATTCATTAATATATTTATCAACATAC 738

QY 271 ACTGTTTCTAAAGTGTCTTTATTTTAACTATTATATGTCAATTTTGAACAGCTCTTTCA 330
Db 737 AATTGATTCATAGCTCTCTCAAGTTTAAAGCTATAGCATCTTGTGTGATATCTTTTTC 678

QY 331 CTTACTAGCAATTTATATTCAGCAACACTTTGTTTATTGTCTAGACTTTTAAAGTTTTCATTC 389
Db 677 ATTAATTCATATACITTTTCATTAAACATCTTGATGGCTAAACAGCAGTAATTTGACTC 619

RESULT 11
US-10-204-708-6
; Sequence 6, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01

```

FILE REFERENCE: DB2a SEQUENCE
; CURRENT APPLICATION NUMBER: US/09/247,352
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 09/026,291
; EARLIER FILING DATE: 1998-02-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 8858
; TYPE: DNA
; ORGANISM: Human and Mouse
US-09-247-352-6

Query Match 5.4%; Score 46.2; DB 4; Length 8858;
Best Local Similarity 53.6%; Pred. No. 0.069;
Matches 191; Conservative 0; Mismatches 456; Indels 1; Gaps 1;

QY 95 TTCCATGAGATGCTCAGTAGATAATCTCTCCATTGTTTAAACATTTGGCTGCTCTCA 154
Db |||||
QY 155 TTTTGACCTATTAAACAGCGCTCTTTAAATATACATCTCTGTAGTGTAGTAGA 214
Db |||||
QY 5919 TTTTATTGTTTTTTTGTGTTATTTTAAATGTTATTTTAAATTTTGTGTTAT 5978
Db |||||
QY 215 ATGAGTGGCTGGAATAAAAGTGGCTGAAT-CATCTCACTAGTAAAGATGCTCAACT 273
Db |||||
QY 5979 TTAATTTGGAATGTTTTTAAATGATGTTATATATATTTTGTAGTTTTTGGG 6038
Db |||||
QY 274 GTTTCCTAAAGTGTATTTTAACTATATATATATATATATTTTGTAGTTTTTGG 633
Db |||||
QY 6039 ATTTTAAATTTTGTAGTATTTTGTATTTTAAATTTTAAATTTTAAAGTTT 6098
Db |||||
QY 334 ACTAGCAATTTATATCAGCAACTGTTTATGTCAGACTTTTAAAGTTTTCATCTG 393
Db |||||
QY 6099 AGTAGTGGTTTTTTAGTTTTTTTTTTTATGAGTGTAGTTATTTTATTTTATTA 6158
Db |||||
QY 394 GNN 453
Db |||||
QY 6159 TTTTGTGTTGTTATATAGTTATTTGTTGTATATATTTTATTTTAAAGTAAATGTA 6218
Db |||||
QY 454 NNN 513
Db |||||
QY 6219 TTTATAGATAGAAATTTGTTTTGTTTATTTGCGATTAGTAAATATAGATATAAG 6278
Db |||||
QY 514 NNN 573
Db |||||
QY 6279 GTACGGTTTTTAGAATTTTGTATGTTTAAAGATATTTAAATTTATTTTATG 6338
Db |||||
QY 574 NNN 633
Db |||||
QY 6339 GAAATAATAAATATGTGAGGAGATGTTTTTATATAGCGTGTTTTATTTTGTGTT 6398
Db |||||
QY 634 ATATTATGTTTGTCTCATTTTAAACAACTGTTTAAATATTTTCAAAATCGACATTATG 693
Db |||||
QY 6399 GTTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 6458
Db |||||
QY 694 ATATATATCTTCCAAAATTTTAAATTTTGTCTTTTTCACATTTT 741
Db |||||
QY 6459 TATAGTTAGGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 6506
Db |||||

RESULT 12
US-09-247-352-6/c
; Sequence 6, Application US/09247352
; Patent No. 6312693
; GENERAL INFORMATION:
; APPLICANT: Aruffo, Alejandro A.
; APPLICANT: Siadak, Anthony W.
; APPLICANT: Berry, Karen K.
; APPLICANT: Harris, Linda
; APPLICANT: Thorne, Barbara A.
; APPLICANT: Bajorath, Jurgen
; APPLICANT: Huse, William D.
; APPLICANT: WJ, Herren
; APPLICANT: Watkins, Jeffery D.
; TITLE OF INVENTION: ANTIBODIES AGAINST HUMAN CD40

FILE REFERENCE: DB2a SEQUENCE
; CURRENT APPLICATION NUMBER: US/09/247,352
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 09/026,291
; EARLIER FILING DATE: 1998-02-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 8858
; TYPE: DNA
; ORGANISM: Human and Mouse
US-09-247-352-6

Query Match 5.4%; Score 46.2; DB 4; Length 8858;
Best Local Similarity 53.6%; Pred. No. 0.069;
Matches 96; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 4 GGTTCGTCATTCACACTTTTAAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGG 63
Db |||||
QY 64 TTTGATGATTTTACATTTATATAGTATTACATTCATGATGATAGTTCTCAGTAGATAATCC 123
Db |||||
QY 3112 CTTGTTAAATTTCTCCTCAATTTGGGTATTTCTAATGATATAAATACACAGTTTAAATACTCA 3053
Db |||||
QY 124 TCCTATTGTTTAAACATTTTGGTGTCTCTCATTTTGGCTTTTGGCTTTTGTAAATAAGCTGCTT 182
Db |||||
QY 3052 TTCTATTGTTGACAACTGGCTATTTCCATTTTGGCTTTTGTAAATAAGCTGCTT 2994
Db |||||

RESULT 13
US-09-466-635-6/c
; Sequence 6, Application US/09466635
; Patent No. 6413514
; GENERAL INFORMATION:
; APPLICANT: Aruffo, Alejandro A.
; APPLICANT: Siadak, Anthony W.
; APPLICANT: Berry, Karen K.
; APPLICANT: Harris, Linda
; APPLICANT: Thorne, Barbara A.
; APPLICANT: Bajorath, Jurgen
; APPLICANT: Thorne, Barbara A.
; TITLE OF INVENTION: ANTIBODIES AGAINST HUMAN CD40
; FILE REFERENCE: DB2 SEQUENCE
; CURRENT APPLICATION NUMBER: US/09/466,635
; CURRENT FILING DATE: 1999-12-17
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 8858
; TYPE: DNA
; ORGANISM: Human and Mouse
US-09-466-635-6

Query Match 5.4%; Score 46.2; DB 4; Length 8858;
Best Local Similarity 53.6%; Pred. No. 0.069;
Matches 96; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 4 GGTTCGTCATTCACACTTTTAAAGTTTATGAGATATATCCATGTTGAATTTTGTAGCTGG 63
Db |||||
QY 3172 GTTTCCTTATGTAACATTTATTTTATGAGATTTCTCTGTCGATGATGATAGTTGTAG 3113
Db |||||
QY 64 TTTGATGATTTTACATTTATATAGTATTACATTCATGATGATAGTTCTCAGTAGATAATCC 123
Db |||||
QY 3112 CTTGTTAAATTTCTCCTCAATTTGGGTATTTCTAATGATATAAATACACAGTTTAAATACTCA 3053
Db |||||
QY 124 TCCTATTGTTTAAACATTTTGGTGTCTCTCATTTTGGCTTTTGTAAATAAGCTGCTT 182
Db |||||
QY 3052 TTCTATTGTTGACAACTGGCTATTTCCATTTTGGCTTTTGTAAATAAGCTGCTT 2994
Db |||||

RESULT 14
US-08-956-171E-521/c
; Sequence 521, Application US/08956171E

Patent No. 6593114
GENERAL INFORMATION: Charles Kunsch
APPLICANT: Gil H. Choi
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Pannou
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER-READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hymen
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 521:
SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 521:
US-08-956-171E-521

Query Match		5.3%;	Score 46;	DB 4;	Length 1353;
Best Local Similarity		26.7%;	Pred. No. 0.052;		
Matches 145;	Conservative 0;	Mismatches 398;	Indels 0;	Gaps 0;	
QY	158	TGACCTATTTTAAACAGCGCCTCTTAAATAATACACTTCCTGTAGTGATGCTAGAAATG	217		
Dd	632	TGTCAGATTATTAATGGAGCCATATTTTATGATTTTCTCTCTGTTTTTCATCTCTA	573		
QY	218	GAGTGGCTGGGAATAAAAAGTGCCTGAATCATCTTCAACTCTAGTAAGATGTCACAATGTT	277		
Dd	572	AAACAGGAATATTATTAGTGCTGTTTTATCTATATCACTAAAAAATTCATCAATATGG	513		
QY	278	TCTAAAAGTGTTTTATTTTAACTATATATATGTCAAATTTTGAACAGCTCTTTCACTACTA	337		
Dd	512	TTTCTAAGTC'TCTTTTTTATTTTCTTAACATCCTTTTGTGTTAACTGTTTTTATTTGCTC	453		
QY	338	GCAATTTATTATCAGCAACACITGTTATGTGTCAGACTTTTAAAGTTTTCATTCACTGNNN	397		
Dd	452	TTTCTTCATCTTTATATCCCATAAATTAACCTACTATATAAACCTTTTCAACATCAT	393		
QY	398	NN	457		
Dd	392	CTCTATAGTTTGATAAAAATGTTTACCCTAAACATTAATCTGATTTTCTAATTTGTTTA	333		
QY	458	NN	517		

Db 332 ATTCTAAGTCGCACTTTAAATTCCTCCTCTTCAAATTAAAGTCCTTCAAATTCAGGGTTATCTT 273
Qy 518 NN 577
Db 272 TTTTTATACTTCAATTTCTTTTAAGAACCTCTCACCTTTGTGTTTTTCAAAGTATTAAATA 213
Qy 578 NNGTGATCATAT 637
Db 212 CACTTTTCATATCTTTTCATGAGATAATCCTAGATAATCTAAATATTACAGTATTTCACATTTT 153
Qy 638 TTATGTTTTTGCATTTTAAAAAACCGTGGTTAAATAATTTCCAAAATCGACATATTGATAT 697
Db 152 TTATTTTTTCTCTTCACTAAAAAATTTCTTCAAAGCCTCATCTKCGCCATATAAATCTCT 93
Qy 698 ATT 700
 ==
Db 92 GTT 90

RESULT 15
US-10-204-708-20
; Sequence 20, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 20
; LENGTH: 6865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-20

Query Match	5.3%;	Score	45.6;	DB	4;	Length	6866;		
Best Local Similarity	31.9%;	Pred. No.	0.09;						
Matches	258;	Conservative	0;	Mismatches	547;	Indels	4;	Gaps	2;
QY	29	ATGAGATATATCCATGTTGAATTTTCTAGCTGTGGTTTG-ATGATTTTACATTATATAG	87						
Db	4867	ATTAGATTTTGTATTTTAGTATATCTTTTTTGTGTTTGTATTTTTTATTATATGAAT	4926						
QY	88	TATTACATTCATGGATAGTCTCAGTGATATCTCTCTATTGTTTAAACATTTTGGCTTG	147						
Db	4927	TTATATTGTTTATAATAAAAAATGAAGAAAAATTTTGTATTGTATATAAATTT	4986						
QY	148	CTTCATATTTTGACCTATTTTAAACAGGCCTCTTTAAATATACACTCTCTGTAGTGTAT	207						
Db	4987	TTATTAGTATTAATAAATATATTTTGTGAATTTTAGATTTTAAATGTGTGTGGGTAA	5046						
QY	208	GCTAGAAATGGAGTGGCTGGAATATAAAGTGGCTGAATCATCTTCAACTCTAGTAGAATGT	267						
Db	5047	TATTTAATGTTAATGTTGGTATTTGGGGAAGTAGATTTTATAAATGTGATAGGTTGT	5106						
QY	268	CAAACTCTTTCTTAAAGTGTTTTATTTTAACTATTATATCTCAATTTTGAACAGCTCTT	327						

```
Db 5107 GGTATATATTTTATTTTGTGTTGGTAAAGTTTTTAATTTGAGGTGAAGATTTTT 5166
QY 328 TCACCTACTAGCAATTTATTATACAGACACTTGTATTGTGACACCTTTTAAGTTTTCAT 387
Db 5167 TTTTGTGTTTGTGTTTGTGAGTAGGTTTATTTT--GTCGTTTAGGTTGGAGT 5223
QY 388 TCACTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 447
Db 5224 GTAGTGGTATCAATTTGGTTATTGTAAATTCGGTTTTTTGGGTTTAAATAATTTTCGTG 5283
QY 448 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 507
Db 5284 TTTTAGTTTTTGTAGTGTGGGATTATAGGTATATATATATATGTGTAGTTAATTTGT 5343
QY 508 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 567
Db 5344 ATTATTAGTAGAGTTGGGTTTCGTTATGTTGGTTAGGTTGGTTTTGAAATTTTGATTTT 5403
QY 568 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 627
Db 5404 AAGTAAATTTGTTTGTGTTTGTGTTTAAAGTGTGAGATTATAGGTATGAGTTATTATT 5463
QY 628 GTGATCATATTTATGTTTGTCTCATTTTAAACAAACCTGTTTAAATATTTTCAACAAATCGAC 687
Db 5464 TTAGTTAAAGTTTGTGTTTGTGTTTATATTTTATGTTTGTGTTTGTGTTTGTGTTTAAAT 5523
QY 688 ATTAGATATATATCTCCAAAATTTTAAATTTTGTCTTTTTTCAATTTTAGTCCTT 747
Db 5524 TATTTTAAATGAGGATTTTAAATTTATTTTATAGAAATTTAAATTTTATTTTATTTT 5583
QY 748 TAGCTCAGCTGGAATTCATTTCTGTGTTGGTGTGAGATAAGTCTTTTCATGTTTTC 807
Db 5584 TTAATTTGGTTTTTGTGTTTGTGTTTATTTTATAGTTTATTTTGTGTTT 5643
QY 808 CTATCAATAAATTTATTCCTTCTGTATT 836
Db 5644 TATAGAGCTTTTTCGATTTTATGTTTATT 5672
```

Search completed: August 14, 2004, 20:48:09
Job time : 85.9109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 2442.69 Seconds
(without alignments)
9989.844 Million cell updates/sec

Title: US-10-082-830-99
Perfect score: 563
Sequence: 1 ggtacatcttggtgtggtat.....gcgcgtagctctgagctctag 563

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vir.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	511.4	90.8	185566	9	AC004168	AC004168 Homo sapi
C 2	53.2	9.4	7218	6	I66494	I66494 Sequence 14
C 3	49.6	8.8	128683	2	AC135793	AC135793 Oryza sat
C 4	48.6	8.6	148526	10	AL732503	AL732503 Mouse DNA
C 5	48.2	8.6	94391	2	AC135575	AC135575 Rattus no
C 6	48.2	8.6	168088	2	AC110880	AC110880 Mus muscu
C 7	48.2	8.6	168773	10	AL731820	AL731820 Mouse DNA
C 8	47.8	8.5	137648	2	AC024316	AC024316 Homo sapi
C 9	47.8	8.5	158608	9	AC051635	AC051635 Homo sapi
C 10	47.8	8.5	193978	2	AC068678	AC068678 Homo sapi
C 11	47.8	8.5	209462	2	AC011814	AC011814 Homo sapi
C 12	47.6	8.5	207777	2	AC102621	AC102621 Mus muscu
C 13	47.6	8.5	222906	2	AC132603	AC132603 Mus muscu
C 14	47.4	8.4	161021	10	EX537253	EX537253 Mouse DNA
C 15	47	8.3	220394	2	AC133842	AC133842 Rattus no
C 16	46.8	8.3	214285	10	AC107635	AC107635 Mus muscu
C 17	46.8	8.3	222845	10	AC100491	AC100491 Mus muscu
C 18	46.6	8.3	65772	2	AC123762	AC123762 Mus muscu
C 19	46.6	8.3	120507	10	AL928895	AL928895 Mouse DNA
C 20	46.6	8.3	177955	2	AC134460	AC134460 Mus muscu
C 21	46.6	8.3	178455	2	AC137969	AC137969 Mus muscu
C 22	46.6	8.3	198829	10	AC120548	AC120548 Mus muscu
C 23	46.6	8.3	200487	2	AC145577	AC145577 Mus muscu
C 24	46.6	8.3	221362	2	AC134084	AC134084 Rattus no
C 25	46.6	8.3	236639	2	AC103213	AC103213 Rattus no
C 26	46.6	8.3	245711	2	AC098021	AC098021 Rattus no
C 27	46.6	8.3	253401	2	AC097127	AC097127 Rattus no
C 28	46.6	8.3	268574	2	AC087129	AC087129 Mus muscu
C 29	46.6	8.3	282426	2	AC145569	AC145569 Mus muscu
C 30	46.2	8.2	157322	9	AC012103	AC012103 Homo sapi
C 31	46.2	8.2	173324	2	AC016038	AC016038 Homo sapi
C 32	46.2	8.2	187557	2	AC069032	AC069032 Homo sapi
C 33	46	8.2	168580	10	AC127696	AC127696 Mus muscu
C 34	46	8.2	179455	2	AC102304	AC102304 Mus muscu
C 35	46	8.2	180017	2	AC115163	AC115163 Rattus no
C 36	46	8.2	191312	10	AL732420	AL732420 Mouse DNA
C 37	46	8.2	201923	10	AC121772	AC121772 Mus muscu
C 38	46	8.2	202010	2	AC136558	AC136558 Rattus no
C 39	46	8.2	202342	10	AL671867	AL671867 Mouse DNA
C 40	46	8.2	209438	2	AC142275	AC142275 Mus muscu
C 41	46	8.2	231704	10	AC051638	AC051638 Mus muscu
C 42	46	8.2	239439	10	AC117573	AC117573 Mus muscu
C 43	46	8.2	243425	2	AC108643	AC108643 Rattus no
C 44	46	8.2	255563	2	AC125834	AC125834 Rattus no
C 45	46	8.2	302176	2	AC096479	AC096479 Rattus no

ALIGNMENTS

RESULT 1
AC004168/c
LOCUS AC004168 Homo sapiens chromosome 4 clone B159L21 map 4q25, complete
DEFINITION sequence.
AC004168
VERSION AC004168.2 GI:4454440
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185566)
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 185566)
AUTHORS      Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE        Direct Submission
JOURNAL      Submitted (23-FEB-1998) Department of Genetics, Stanford Human
              Genome Center, 955 California Avenue, Palo Alto, CA 94304, USA
REFERENCE    3 (bases 1 to 185566)
AUTHORS      Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
TITLE        Direct Submission
JOURNAL      Submitted (19-MAR-1999) Department of Genetics, Stanford Human
              Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA
COMMENT      On Mar 19, 1999 this sequence version replaced gi:3603082.
              Quality: Phrap Quality >=40 100.0% of Sequence;
              Estimated Total Number of Errors is 0.0.
              STS Content:
              SHGC-50970 G33969
              SHGC-23829 G28472
              WI-30731 G23237
              SHGC-23845 G28477
              SHGC-23843 G28476
              SHGC-50971 G34439.
FEATURES     Location/Qualifiers
              1..185566
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /chromosome="4"
               /map="4q25"
               /clone="B159121"
               /clone_lib="Caltech/Research Genetics Whole-Genome BAC
               Library"
ORIGIN
Query Match      90.8%; Score 511.4; DB 9; Length 185566;
Best Local Similarity 93.6%; Pred. No. 6.3e-109;
Matches 523; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 GGTACATCTTGGCTGTGGATGGAAATTTGACATACATCTTTTATTTTCTTTACCTGTTACAT 60
Db 25328 GGTACATCTTGGCTGTGGATGGAAATTTGACATACATCTTTTATTTTCTTTACCTGTTACAT 25269
QY 61 ATCAAACTTTAGGATGATTACTTCCAAAGCGGTTAACTTATTCAGATTTGACTGAGTC 120
Db 25268 ATCAAACTTTAGGATGATTACTTCCAAAGCGGTTAACTTATTCAGATTTGACTGAGTC 25209
QY 121 TCCTATTTTCTTAAATTTACTAGAGTGAGGCTCCAGAACTACAGAAATAGAGGAA 180
Db 25208 TCCTATTTTCTTAAATTTACTAGAGTGAGGCTCCAGAACTACAGAAATAGAGGAA 25149
QY 181 AGTCTCCATTGAGCCATGAATCTGAGCACCTGGCATTAAAGCATGAAGATGAGGCTTC 240
Db 25148 AGTCTCCATTGAGCCATGAATCTGAGCACCTGGCATTAAAGCATGAAGATGAGGCTTC 25089
QY 241 TATGGTAGGACTGGAGTAGGAGCATTCAGAAAGGATCTCAGAGGTCAGAAACAATA 300
Db 25088 TATGGTAGGACTGGAGTAGGAGCATTCAGAAAGGATCTCAGAGGTCAGAAACAATA 25029
QY 301 GATTATCAGTTAAATATCTCTGGACCAAGAAAGACCTTGAATCCCTGGCTCGGTGATGA 360
Db 25028 GATTATCAGTTAAATATCTCTGGACCAAGAAAGACCTTGAATCCCTGGCTCGGTGATGA 24969
QY 361 TACATATTTAATACAAATGAGAGCTCTTGGAGTGAGTGAGAGTATAGGGCTGAAG 420
Db 24968 TACATATTTAATACAAATGAGAGCTCTTGGAGTGAGTGAGAGTATAGGGCTGAAG 24909
QY 421 AACAGGAGACAGAAGCAATTCCTTACACAGAAAGCTGATTATATAGATAATG 480
Db 24908 AACAGGAGACAGAAGCAATTCCTTACACAGAAAGCTGATTATATAGATAATG 24849
QY 481 GTACATTCATATAAATCATGATATTTACTATTAATAAATAACCA 525
Db 24848 GTACATTCATAT-AACATGATATATTTACTATTAATAAATAACCA 24805

```

```

RESULT 2
LOCUS       I66494
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1 GI:2724471
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 7218)
AUTHORS     Dorrer,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE       Recombinant Fowlpox virus
JOURNAL     Patent: US 5670367-A 14 23-SEP-1997;
FEATURES     Location/Qualifiers
              source
               1..7218
               /organism="unknown"
               /mol_type="unassigned DNA"
ORIGIN
Query Match      9.4%; Score 53.2; DB 6; Length 7218;
Best Local Similarity 1.9%; Pred. No. 0.033;
Matches 7; Conservative 218; Mismatches 141; Indels 0; Gaps 0;
QY 135 AATTACTAGAAAGTGAGGCTCCAAGAACTACAGAAATAGAGAAAGTCTTCATTGAGC 194
Db 1444 AATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1385
QY 195 CATGAACCTGTGACACCTGGCATTTAAGCATGAAGATGAGGCTTCATGTTAGGACTG 254
Db 1384 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1325
QY 255 GAGTAGGCAGCATTCAGGAAAGATCTCAGAGGTCAGAAACATAGATATTTCAGTTAAA 314
Db 1324 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1265
QY 315 TACTTCGGACCAAGAGACCTTGAATCTGCTCGGTGATGATATACATATTATATAC 374
Db 1264 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1205
QY 375 ACAATGAGAACTCCTTGAGGCTGAGTGAGTATGAGGCTCAAGAACAGGACAGAA 434
Db 1204 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1145
QY 435 GACAACTCAATGCTCTTACAGAAAGCTGATTAATATAGATATGATGATCATCATATA 494
Db 1144 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1085
QY 495 ACATGA 500
Db 1084 RRRRR 1079

RESULT 3
LOCUS       AC135793
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 11 clone
ACCESSION   AC135793
VERSION     AC135793.3 GI:38153795
KEYWORDS    HTG; HTGS PHASE2; HTGS ACTIVEFIN.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzaceae; Oryza.
              1 (bases 1 to 128683)
              Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M.,
              Overton II,L., Tsitrin,T., Kim,M., Bera,J., Jin,S., Fadrosch,D.W.,
              Tallon,L., Koo,H., Zismann,V., Hsiao,J., Blunt,S., Vanaken,S.,
              Riedmuller,S.B., Utterbach,T., Feldblyum,T., Yang,Q., Haas,B.,
              Suh,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and

```



```

Fraser, C.
Oryza sativa ssp. japonica cv. Nipponbare TIGR CLONE BAC genomic
sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128683)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
REFERENCE 3 (bases 1 to 128683)
AUTHORS Buell, R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
COMMENT
* NOTE: This is a "working draft" sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 128683: contig of 128683 bp in length.
FEATURES
Location/Qualifiers
1..128683
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="11"
/clone="OSJNB0017P15"
/note="japonica cultivar-group"
ORIGIN
Query Match 8.8%; Score 49.6; DB 2; Length 128683;
Best Local Similarity 11.9%; Pred. No. 0.14; Mismatches 232; Indels 4; Gaps 2;
Matches 60; Conservative 232;
32 ATACTTTTATTTCTTACCTGTATACATATCAATCTTAGAGTGTATTCTTCAAGGC 91
50753 RMYRKTCTKWKYKSTTTGCAAMWTAGKMYGKMTWTGSCYKRYWRAWYCYKWSMK 50694
QY 92 GGTAAACTTATTCAGATTACACTGAGTCTCTTATTTCTTAAATTTACTAGAGTGAG 151
DB 50693 RMKSYATRYMYRSWYTSWTSYMKMYRSGWNCWYMYRKMAYWYTMKWKWGRGA 50634
QY 152 GCTCAAGAACTACAGAAATAGAGAAAGTCTCTTATGAGCCATGAACTGTGACACC 211
DB 50633 MMWYWCRRMMAWMYRCWTXGTGASWKNMSASAAKAWYCRMKRKKWASYTRMK 50574
QY 212 TGGCATTTAAGCATGAGAGTAGGGCTTCTATGTTAGGAGTGGAGTAGCAGCATCCA 271
DB 50573 WKRWARWATYRGRARWYTMKMBWYTKOMARRAKRWAYRARAASSSMARRTTY-- 50516
QY 272 GGAAGAGTCTCAGAGGTCAAGAAACAATAGATTATCAGTTAAATATCTTCTGGACCAAAGA 331
DB 50515 -RMSASSRKKWMMWMSWRSRYWMSWMSMKRSYRKWKYAKCAIYEMWWSYVYASAT 50457
QY 332 AGACCTTGAATCTTGGCTCGGTGATGATATACATATTTAATACATGAGAGTCCCTT 391
DB 50456 GKWMAGKTAKRSMYTRSAWMTYWSWKKWYWGATKRYWTKWAKRAWWTKTATY 50397
QY 392 GAGGTGAGTGAAGTATAGAGGTGAGAGCAACAGACAGACAGACAATTTCAATGTCTCT 451
DB 50396 YCRYATTATWTTCTTAGSMMW-WWYTRKCMCRMTWAKWSYATRSYKTKRSKSKKRYK 50338
QY 452 TACACAGAAAGTGAATATATAGATATATGTTACATTCATATAACATGATATATTACTA 511
DB 50337 RYRMKRGKYARKWRSRRRKKWWTGYGWCRRYWSYWRMTAMSAWMTAAWMYRKA 50278
QY 512 ATTAACAACATCAACACACACAA 535

```

```

Db 50277 MWCCARCAATAAKASYAKREWTW 50254
RESULT 4
AL732503 148526 bp DNA linear ROD 25-JUL-2002
LOCUS Mouse DNA sequence from clone RP23-2809 on chromosome 4, complete
DEFINITION sequence.
ACCESSION AL732503
VERSION AL732503.8 GI:21998252
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 148526)
AUTHORS Pearce, A.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquy@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 26, 2002 this sequence version replaced gi:21911625.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquy@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TRMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-2809 is
from the RPI-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES
Location/Qualifiers
1..148526
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-2809"
/clone_lib="RPI-23"
ORIGIN
Query Match 8.6%; Score 48.6; DB 10; Length 148526;
Best Local Similarity 64.9%; Pred. No. 0.23;
Matches 72; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 427 AGACAGAGCAATTCCTTACACAGAGTCTTATATAGATATAGTATGTTACAT 486
DB 135971 AGCCAGAAAAATCCAGATGCTCTTCAACAGAGGATGTGTACAGAAATGTTGATCAT 136030
QY 487 TCATATAAATCATGATATTTTCTAATTAATAAATCAACACACACACA 537
DB 136031 TTACACAATGGAGTATTCTTCTGCTATTAAAAAATGAATTCATGAACA 136081

```

RESULT 5
 AC135575 94391 bp DNA linear HTG 19-OCT-2002
 LOCUS Rattus norvegicus clone CH230-20316, *** SEQUENCING IN PROGRESS
 DEFINITION *** 52 unordered pieces.
 AC135575
 ACCESSION AC135575
 VERSION HTG: HTGS PHASE1
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 94391)
 Muny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Auguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gbregregoris, E., Geer, K., Gill, R., Grady, N., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulvik, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensueta, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Margum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mathiney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, N., Mundaas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwokeneme, O., Okunolu, G.,
 Olarunpungoon, A., Pal, S., Parks, K., Pastermak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rokey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,
 Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wiczek, R.,
 Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 94391)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-OCT-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine

Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KCOF
 Center clone name: CH230-20316
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 38760 bases at least Q40
 Consensus quality: 41115 bases at least Q30
 Consensus quality: 42555 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 52 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1118: contig of 1118 bp in length
 1119 1218: gap of unknown length
 1219 2501: contig of 1283 bp in length
 2502 2601: gap of unknown length
 2602 3811: contig of 1216 bp in length
 3818 3917: gap of unknown length
 3918 5144: contig of 1227 bp in length
 5145 5244: gap of unknown length
 5245 6260: contig of 1016 bp in length
 6261 6360: gap of unknown length
 6361 7588: contig of 1225 bp in length
 7589 8694: gap of unknown length
 8695 10206: contig of 1412 bp in length
 10207 10306: gap of unknown length
 10307 11520: contig of 1214 bp in length
 11521 11620: gap of unknown length
 11621 13019: contig of 1399 bp in length
 13020 13119: gap of unknown length
 13120 14579: contig of 1460 bp in length
 14580 14679: gap of unknown length
 14680 16092: contig of 1413 bp in length
 16093 16192: gap of unknown length
 16193 17509: contig of 1317 bp in length
 17510 17609: gap of unknown length
 17610 18996: contig of 1387 bp in length
 18997 19096: gap of unknown length
 19097 20159: contig of 1063 bp in length
 20160 20259: gap of unknown length
 20260 21714: contig of 1455 bp in length
 21715 21844: gap of unknown length
 21845 23289: contig of 1475 bp in length
 23290 23389: gap of unknown length
 23390 25099: contig of 1710 bp in length
 25100 25199: gap of unknown length
 25200 26334: contig of 1135 bp in length
 26335 26434: gap of unknown length
 26435 27867: contig of 1433 bp in length
 27868 27967: gap of unknown length
 27968 29046: contig of 1079 bp in length
 29047 29146: gap of unknown length
 29147 30488: contig of 1342 bp in length
 30489 32559: contig of 1671 bp in length
 32560 32559: gap of unknown length
 32560 33897: contig of 1538 bp in length
 33898 33997: gap of unknown length
 33998 35361: contig of 1364 bp in length

```

* 35362 35461: gap of unknown length
* 35462 36914: contig of 1453 bp in length
* 36915 37014: gap of unknown length
* 37015 38663: contig of 1649 bp in length
* 38664 38763: gap of unknown length
* 38764 39847: contig of 1084 bp in length
* 39848 39947: gap of unknown length
* 39948 41838: contig of 1891 bp in length
* 41839 41938: gap of unknown length
* 41939 44158: contig of 2220 bp in length
* 44159 44258: gap of unknown length
* 44259 45586: contig of 1328 bp in length
* 45587 45686: gap of unknown length
* 45687 47721: contig of 2035 bp in length
* 47722 47821: gap of unknown length
* 47822 49216: contig of 1395 bp in length
* 49217 49316: gap of unknown length
* 49317 51012: contig of 1696 bp in length
* 51013 51112: gap of unknown length
* 51113 53593: contig of 2481 bp in length
* 53594 53693: gap of unknown length
* 53694 55129: contig of 1436 bp in length
* 55130 55229: gap of unknown length
* 55230 56905: contig of 1676 bp in length
* 56906 57005: gap of unknown length
* 57006 58201: contig of 1196 bp in length
* 58202 58301: gap of unknown length
* 58302 60358: contig of 1957 bp in length
* 60359 61875: contig of 1517 bp in length
* 61876 61975: gap of unknown length
* 61976 63220: contig of 1145 bp in length
* 63221 65269: contig of 2049 bp in length
* 65270 65369: gap of unknown length
* 65370 67808: contig of 2439 bp in length
* 67809 67908: gap of unknown length
* 67909 71132: contig of 3224 bp in length
* 71133 71233: gap of unknown length
* 71234 73404: contig of 2172 bp in length
* 73405 73505: gap of unknown length
* 73506 76370: contig of 2866 bp in length
* 76371 76470: gap of unknown length
* 76471 78273: contig of 1803 bp in length
* 78274 78373: gap of unknown length
* 78374 82072: contig of 3699 bp in length
* 82073 82172: gap of unknown length
* 82173 84747: contig of 2575 bp in length
* 84748 84848: gap of unknown length
* 84849 87309: contig of 2462 bp in length
* 87310 87409: gap of unknown length
* 87410 89644: contig of 2235 bp in length
* 89645 89744: gap of unknown length
* 89745 94391: contig of 4647 bp in length.
FEATURES
    source
    1..94391
    /organism="Rattus norvegicus"

Query Match      8.6%; Score 48.2; DB 2; Length 94391;
Best Local Similarity 61.6%; Pred. No. 0.31;
Matches 77; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACAAATTCAAATGCTTACACAGAACTGATTATAGATA 477
DB 31033 AATGACCAAGAGCTGGAAAGAAATGTTGATGTCCTTCAACAGAGGAAAGAAATATAGAAAC 31092
QY 478 ATGGTACATTCATATACATCATATATTTACTATTAATTAACATACACACACACACA 537
DB 31093 GTGGTACATTCATATACATCATATTTACTATTAATTAACATACATTCGCGAATGAATT 31152
QY 538 CCTCG 542
DB 31153 CCTAG 31157

```

RESULT 6 AC110880 LOCUS

DEFINITION

AC110880.4 GI:24211407
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)

ACCESSION

1 (bases 1 to 168088)
Birren,B., Nusbaum,C. and Lander,E.

REFERENCE

1 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

2 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

JOURNAL

3 (bases 1 to 168088)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

AUTHORS

Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhtalter,B.,

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L18747
 Center clone name: 456.G.3

----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 166916 bases at least Q40
 Consensus quality: 167532 bases at least Q30
 Consensus quality: 167776 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 167889; sum-of-contigs
 Quality coverage: 12.3 in Q20 bases; agarose-fp
 Quality coverage: 12.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 43630: contig of 43630 bp in length
 * 43631 43730: gap of 100 bp
 * 43731 83102: contig of 39372 bp in length
 * 83103 83202: gap of 100 bp
 * 83203 168088: contig of 84886 bp in length.

FEATURES

source
 1. 168088
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP24-456G3"
 /clone_lib="RPCI-24 Male Mouse BAC"
 misc_feature
 1. 43630
 /note="assembly_fragment"
 misc_feature
 43731..83102
 /note="assembly_fragment"
 misc_feature
 83203..168088
 /note="assembly_fragment"

ORIGIN

Query Match 8.6%; Score 48.2; DB 2; Length 168088;
 Best Local Similarity 61.6%; Pred. No. 0.28; Indels 0; Gaps 0;
 Matches 77; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
 QY 418 AAGAACAGGACAGAGACAATTCCTTACACAGAGACTGATTATATAGATA 477
 Db 162492 AATAGCCAGAGCTGGAAAGACCAGATGCTCCTCAACAGAGGATGGATACAGAAAT 162551
 QY 478 ATGTACATTCATATAACATGATATATTTACTAATTAACATACACACACACA 537
 Db 162552 ATGTACATTCATATAACATGATATATTTACTAATTAACATACACACACA 537
 QY 538 CCTCG 542
 Db 162612 CTTAG 162616

RESULT 7
 AL731820
 LOCUS AL731820 168773 bp DNA linear ROD 26-JUL-2002
 DEFINITION Mouse DNA sequence from clone RP23-135H19 on chromosome X, complete
 sequence.
 ACCESSION AL731820
 VERSION AL731820
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 168773)
 Griffiths, C.
 Direct Submission
 Submitted (26-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jul 29, 2002 this sequence version replaced gi:21955608.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW:
 SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-135H19 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6.

FEATURES

source
 1. 168773
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-135H19"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 8.6%; Score 48.2; DB 10; Length 168773;
 Best Local Similarity 67.3%; Pred. No. 0.28;
 Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 418 AAGAACAGGACAGAGACAATTCCTTACACAGAGACTGATTATATAGATA 477
 Db 82026 AATAGCCAGAACTGGAAAGATCCAGATGCTCCTTCAACAGAGATGGATACAGAAAT 82085
 QY 478 ATGTACATTCATATAACATGATATATTTACTAATTAATAAAA 518
 Db 82086 GTGGTACATTTATATAAGGAGTCTTATTCAGCTATTAAAA 82126

RESULT 8

AC024316/c
 LOCUS AC024316 137648 bp DNA linear HTG 26-SEP-2000
 DEFINITION Homo sapiens clone RP11-2608, WORKING DRAFT SEQUENCE, 17 unordered
 pieces.
 ACCESSION AC024316
 VERSION AC024316.3 GI:10305255
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 137648)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-2608
Unpublished
2 (bases 1 to 137648)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,
Bouckgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepel, X., Collangelo, M., Collins, S., Collamore, A., Cooke, P.,
DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M.,
Fenesor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., Lagocque, K., Lehoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McCarthy, M.,
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Olivari, T. M.,
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 26, 2000 this sequence version replaced gi:10198462.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4672
Center clone name: 26_O8
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 128694 bases at least Q40
Consensus quality: 133079 bases at least Q30
Consensus quality: 135006 bases at least Q20
Insert size: 137000; agarose-fp
Insert size: 136048; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 3290: contig of 3290 bp in length
* 3291 3390: gap of 100 bp
* 3391 5203: contig of 1813 bp in length
* 5204 5304: gap of 100 bp
* 5304 7364: contig of 2061 bp in length
* 7365 7465: gap of 100 bp
* 7466 10939: contig of 3475 bp in length
* 10940 11039: gap of 100 bp
* 11040 15415: contig of 4376 bp in length
* 15416 15515: gap of 100 bp
* 15516 20495: contig of 4980 bp in length
* 20496 20595: gap of 100 bp
* 20596 23962: contig of 5367 bp in length

```

```

* 25963 26062: gap of 100 bp
* 26063 30774: contig of 4712 bp in length
* 30775 30874: gap of 100 bp
* 30875 36005: contig of 5131 bp in length
* 36006 36105: gap of 100 bp
* 36106 40573: contig of 4468 bp in length
* 40574 40674: gap of 100 bp
* 40675 47316: contig of 6643 bp in length
* 47317 47416: gap of 100 bp
* 47417 57551: contig of 10135 bp in length
* 57552 57651: gap of 100 bp
* 57652 67679: contig of 10028 bp in length
* 67680 67779: gap of 100 bp
* 67780 78221: contig of 10342 bp in length
* 78222 105075: contig of 26854 bp in length
* 105076 105175: gap of 100 bp
* 105176 124687: contig of 19512 bp in length
* 124688 124788: gap of 100 bp
* 124789 137648: contig of 12861 bp in length.

```

FEATURES
source

```

1..137648
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="RP11-2608"
/clone="RP11-2608"

```

```

1..3290
/note="assembly_fragment"
vector_side:left

```

```

misc_feature
1..3290
/note="assembly_fragment"
vector_side:left

```

```

misc_feature
3391..5203
/note="assembly_fragment"

```

```

misc_feature
5304..7364
/note="assembly_fragment"

```

```

misc_feature
7465..10939
/note="assembly_fragment"

```

```

misc_feature
11040..15415
/note="assembly_fragment"

```

```

misc_feature
15516..20495
/note="assembly_fragment"

```

```

misc_feature
20596..25962
/note="assembly_fragment"

```

```

misc_feature
26063..30774
/note="assembly_fragment"

```

```

misc_feature
30875..36005
/note="assembly_fragment"

```

```

misc_feature
36106..40573
/note="assembly_fragment"

```

```

misc_feature
40674..47316
/note="assembly_fragment"

```

```

misc_feature
47417..57551
/note="assembly_fragment"

```

```

misc_feature
57652..67679
/note="assembly_fragment"

```

```

misc_feature
67780..78221
/note="assembly_fragment"

```

```

misc_feature
78222..105075
/note="assembly_fragment"

```

```

misc_feature
105176..124687
/note="assembly_fragment"

```

```

misc_feature
124788..137648
/note="assembly_fragment"

```

```

vector_side:right

```

ORIGIN

```

Query Match      8.5%; Score 47.8; DB 2; Length 137648;
Best Local Similarity 61.8%; Pred. No. 0.36;
Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```

418 AAGACAGGAGACAGACAGACATTCAATGCTCTACAGAGACTGATTATATAGATA 477

Db 116769 AACGCCAGAAAATGGAAACAAATCAAAATGTCCTCATCAACACAGATGGATAAATAGTT 1167710

QY 478 ATGGTACATTCATATAAACAATGATATATATTTACTAAATTAACACATACACACACACAA 537

Db 116709 GTGGTATATTTACACACAGAAAAAATGGCAGTAGAAGAACTTACATTCACACACATCA 116650

QY 538 CCT: 540

Db 116649 CCT 116647

RESULT 9

LOCUS AC051635/c

DEFINITION Homo sapiens chromosome 18, clone RP11-540A4, complete sequence.

ACCESSION AC051635

VERSION AC051635.7 GI:17298634

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

TITLE Homo sapiens chromosome 18, clone RP11-540A4

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Boguslavsky,L., Bouckhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Hahne,W., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 5 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 6 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Dec 4, 2001 this sequence version replaced gi16756307. All repeats were identified using RepeatMasker: Smit, A.P.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5261

Center clone name: 540_A_4

Location/Qualifiers

1. .158608

/organism="Homo sapiens"

/mol_type="genomic DNA"

TITLE

JOURNAL

REFERENCE

AUTHORS

Direct Submission

Submitted (04-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 5 (bases 1 to 158608)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (27-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Dec 4, 2001 this sequence version replaced gi16756307. All repeats were identified using RepeatMasker: Smit, A.P.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5261

Center clone name: 540_A_4

Location/Qualifiers

1. .158608

/organism="Homo sapiens"

/mol_type="genomic DNA"

```

/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-540A4"
/clone_lib="RC1-11 Human Male BAC"
8..106
/rpt_family="AluSg/x"
15..27
/note="<30 qual SNGL region"
75..77
/note="<30 qual SNGL region"
106..109
/note="<30 qual SNGL region"
112..228
/rpt_family="L1MB2"
complement(563..632)
/rpt_family="MIR"
1053..1057
/note="<30 qual SNGL region"
complement(1123..1352)
/rpt_family="MIR"
1624..1702
/rpt_family="L2"
complement(1853..2218)
/rpt_family="L2"
3091..3358
/rpt_family="L1MC5"
3357..3850
/rpt_family="L1MC1"
complement(3868..4006)
/rpt_family="LTR7"
4011..4167
/rpt_family="MIR3"
complement(4235..4560)
/rpt_family="LTR16C"
5345..5372
/rpt_family=" (TTTG)n"
5957..5984
/rpt_family=" (TTTG)n"
6132..6246
/rpt_family="MIR3"
6598..6718
/rpt_family="L1MBd"
6731..6769
/rpt_family=" (CA)n"
6940..7148
/rpt_family="L1MBd"
7153..7486
/rpt_family="L1MB8"
7510..8250
/rpt_family="L1MBd"
8279..8516
/rpt_family="L1MBd"
8612..8827
/rpt_family="AluJb"
9027..9126
/rpt_family="L1MBd"
9127..9435
/rpt_family="AluY"
9436..9486
/rpt_family="L1MBd"
9487..9805
/rpt_family="AluSx"
9806..9922
/rpt_family="L1MBd"
complement(9940..9989)
/rpt_family="MLTII"

Query Match      8.5%; Score 47.8; DB 9; Length 158608;
Best Local Similarity 61.8%; Pred. No. 0.35;
Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

418 AAGAACAGGACAGACAGACAAATGTCCTTACACAGAGACTGATTATATAGATA 477

```

Db 134152 AACAGCCAGAAAATGGAACAATACAAATGTCATCAACACAGAAATGGAATAAAGTT 134093
 QY 478 ATGGTACATTTCATATAAATCATATATTTTACTTAATTTAAACATACACACACACACCA 537
 Db 134092 GTGGTATATTTACACACAGAAAAAATGCGCAGTAGAAGAACTTACATTCACACATCA 134033
 QY 538 CCT 540
 Db 134032 CCT 134030

RESULT 10
 AC068678
 LOCUS
 DEFINITION Homo sapiens chromosome 18 clone RP11-104N14 map 18, WORKING DRAFT
 SEQUENCE, 5 unordered pieces.
 AC068678
 VERSION AC068678.3 GI:16506981
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 193978)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 2 (bases 1 to 193978)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellan,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lechoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thumann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission

TITLE
 JOURNAL
 COMMENT Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141 USA
 On Oct 28, 2001 this sequence version replaced GI:13249473.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIER
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1295
 Center clone name: 104.N.14
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 4% of reads
 Sequencing vector: Plasmid; n/a; 96% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 191018 bases at least Q40
 Consensus quality: 192308 bases at least Q30
 Consensus quality: 192875 bases at least Q20

Insert size: 190000; agarose-fp
 Insert size: 193578; sum-of-ctnigs
 Quality coverage: 7.3 in Q20 bases; agarose-fp
 Quality coverage: 7.1 in Q20 ba.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 ctnigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the ctnigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 59978: contig of 59978 bp in length
 * 59979 60078: gap of 100 bp
 * 60079 83937: contig of 23859 bp in length
 * 83938 84037: gap of 100 bp
 * 84038 113621: contig of 29584 bp in length
 * 113622 113721: gap of 100 bp
 * 113722 147302: contig of 33581 bp in length
 * 147303 147402: gap of 100 bp
 * 147403 193978: contig of 46576 bp in length.

FEATURES

source

Location/Qualifiers
 1. 193978
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-104N14"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature
 1. 59978
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"

misc_feature.
 60079..83937
 /note="assembly_fragment"

misc_feature
 84038..113621
 /note="assembly_fragment"

misc_feature
 113722..147302
 /note="assembly_fragment"

misc_feature
 147403..193978
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 8.5%; Score 47.8; DB 2; Length 193978;
 Best Local Similarity 61.8%; Pred. No. 0.34;
 Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 418 AAGACAGGACAGACAGACAAATGTCCTTACACAGACGACTGATTATATAGATA 477
 DB 101530 AACCCAGAAATGGAAACATACAAATGTCCTTACACAGACGAAATGATTAAGTT 101589
 QY 478 ATGTACATTCATATAACATGATATATTTACTTAATTAACATACACACACACACA 537
 DB 101590 GTGGTATATTTACACACAGAAAAAATGGCAGTAGAAGAACTTACATTCACACACATCA 101649
 QY 538 CCT 540
 DB 101650 CCT 101652

RESULT 11
 AC011814
 LOCUS AC011814 209462 bp DNA linear HTG 26-MAY-2000
 DEFINITION Homo sapiens chromosome 18 clone RP11-153D16 map 18, WORKING DRAFT
 SEQUENCE, 40 unordered pieces.
 ACCESSION AC011814
 VERSION AC011814.3 GI:8072439
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 209462)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 18, clone RP11-153D16
 Unpublished

2 (bases 1 to 209462)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukigalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hago,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karstad,A., Klein,J.,
 Lehotzky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:7321509.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1307

Center clone name: 153_D16

----- Summary Statistics

Sequencing vector: M13; M7815; 97% of reads

Sequencing vector: Plasmid; n/a; %0-f% of reads

3.29835082458771Chemistry: Dye-terminator; Dye-terminator; 8% of reads

Chemistry: Dye-terminator Big Dye; 92% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 172516 bases at least Q40

Consensus quality: 185069 bases at least Q30

Consensus quality: 195901 bases at least Q20

Insert size: 175000; agarose-fp

Insert size: 205562; sum-of-ctnigs

Quality coverage

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 40 ctnigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the ctnigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1353: contig of 1353 bp in length

1354 1453: gap of 100 bp

1454 2778: contig of 1325 bp in length

2779 2878: gap of 100 bp

2879 4211: contig of 1333 bp in length

4212 4311: gap of 100 bp

4312 5622: contig of 1311 bp in length

5623 5722: gap of 100 bp

5723 7256: contig of 1534 bp in length

7257 7357: gap of 100 bp

7358 8828: contig of 1472 bp in length

8829 10225: contig of 1297 bp in length

10226 10326: gap of 100 bp

10327 11376: contig of 1051 bp in length

11377 11476: gap of 100 bp

11477 12643: contig of 1167 bp in length

* 12644 12743: gap of 100 bp
* 12744 13857: contig of 1114 bp in length
* 13858 13957: gap of 100 bp
* 13958 15355: contig of 1398 bp in length
* 15356 15455: gap of 100 bp
* 15456 17127: contig of 1672 bp in length
* 17128 17227: gap of 100 bp
* 17228 18117: contig of 1890 bp in length
* 18118 18217: gap of 100 bp
* 18218 20984: contig of 1767 bp in length
* 20985 21084: gap of 100 bp
* 21085 23101: contig of 2017 bp in length
* 23102 23201: gap of 100 bp
* 23202 25331: contig of 2130 bp in length
* 25332 25431: gap of 100 bp
* 25432 27214: contig of 1783 bp in length
* 27215 27314: gap of 100 bp
* 27315 29160: contig of 1846 bp in length
* 29161 29260: gap of 100 bp
* 29261 32570: contig of 3310 bp in length
* 32571 32670: gap of 100 bp
* 32671 34880: contig of 2210 bp in length
* 34881 34981: gap of 100 bp
* 34982 38325: contig of 3345 bp in length
* 38326 38425: gap of 100 bp
* 38426 42527: contig of 4102 bp in length
* 42528 42627: gap of 100 bp
* 42628 46099: contig of 3472 bp in length
* 46100 48199: gap of 100 bp
* 48200 49618: contig of 3419 bp in length
* 49619 53310: contig of 3592 bp in length
* 53311 53410: gap of 100 bp
* 53411 56872: contig of 3362 bp in length
* 56873 60358: gap of 100 bp
* 60359 60458: contig of 3486 bp in length
* 60459 66617: contig of 6159 bp in length
* 66618 66717: gap of 100 bp
* 66718 75403: contig of 8686 bp in length
* 75404 75503: gap of 100 bp
* 75504 83115: contig of 7612 bp in length
* 83116 83215: gap of 100 bp
* 83216 91822: contig of 8607 bp in length
* 91823 91922: gap of 100 bp
* 91923 101265: contig of 9343 bp in length
* 101266 101365: gap of 100 bp
* 101366 111447: contig of 10082 bp in length
* 111448 111547: gap of 100 bp
* 111548 121183: contig of 9636 bp in length
* 121184 133655: contig of 12372 bp in length
* 133656 133755: gap of 100 bp
* 133756 147893: contig of 14138 bp in length
* 147894 147993: gap of 100 bp
* 147994 160302: contig of 12309 bp in length
* 160303 160402: gap of 100 bp
* 160403 174936: contig of 14534 bp in length
* 174937 175036: gap of 100 bp
* 175037 189972: contig of 14936 bp in length
* 189973 190072: gap of 100 bp
* 190073 209462: contig of 19390 bp in length.

FEATURES
source

1..209462
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-153D16"
/clone_lib="RPC1-11 Human Male BAC"
1..1353
/note="assembly_fragment"

misc_feature

misc_feature 1454..2778
/note="assembly_fragment"
misc_feature 2879..4211
/note="assembly_fragment"
misc_feature 4312..5622
/note="assembly_fragment"
misc_feature 5723..7256
/note="assembly_fragment"
misc_feature 7357..8928
/note="assembly_fragment"
misc_feature 8929..10225
/note="assembly_fragment"
misc_feature 10326..11376
/note="assembly_fragment"
misc_feature 11477..12643
/note="assembly_fragment"
misc_feature 12744..13857
/note="assembly_fragment"
misc_feature 13958..15355
/note="assembly_fragment"
misc_feature 15456..17127
/note="assembly_fragment"
misc_feature 17228..18117
/note="assembly_fragment"
misc_feature 18218..20984
/note="assembly_fragment"
misc_feature 21085..23101
/note="assembly_fragment"
misc_feature 23202..25331
/note="assembly_fragment"
misc_feature 25432..27214
/note="assembly_fragment"
misc_feature 27315..29160
/note="assembly_fragment"
misc_feature 29261..32570
/note="assembly_fragment"
misc_feature 32671..34880
/note="assembly_fragment"
misc_feature 34981..38325
/note="assembly_fragment"
misc_feature 38426..42527
/note="assembly_fragment"
misc_feature 42628..46099
/note="assembly_fragment"
misc_feature 46200..49618
/note="assembly_fragment"

Query Match 8.5%; Score 47.8; DB 2; Length 209462;
Best Local Similarity 61.8%; Pred.No. 0.34;
Matches 76; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 418 AAGAACAGGAGAGAGAACAAATGCTTACACAGAACTGATTATATAGATA 477
Db 65547 AACAGCCAGAAATGGAACATACATGTCATCAACACAGATGATAAATAGTT 65606
QY 478 ATGGTACATTATATAACATGATATATTACTAATTAACATACACACACACA 537
Db 65607 GTGGTATATTTACACACAGAAAAAATGCGTAGAGAAAGACTTACATTCACACACATCA 65666
QY 538 CCT 540
Db 65667 CCT 65669

RESULT 12
AC102621/c
LOCUS AC102621 207777 bp DNA linear HTG 16-OCT-2003
DEFINITION Mus musculus chromosome 1 clone RP23-42968 map 1, *** SEQUENCING IN
PROGRESS ***, 5 unordered pieces.
ACCESSION AC102621 GI:37693693
VERSION AC102621.4 HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)

Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@watson.wustl.edu

----- Project Information -----
 Center project name: M_BB066C10

----- Summary Statistics -----

Sequencing vector: M13; 0%
 Chemistry: Dye-terminator; Plasmid; 100%
 Chemistry: Dye-terminator; Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 315826 bases at least Q40
 Consensus quality: 317232 bases at least Q30
 Consensus quality: 317973 bases at least Q20
 Insert size: 223000; agarose-fp
 Insert size: 322006; sum-of-contigs
 Quality coverage: 13.22 in Q20 bases; agarose-fp
 Quality coverage: 8.88 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1483: contig of 1483 bp in length
 * 1484 1583: gap of unknown length
 * 1584 3319: contig of 1736 bp in length
 * 3320 3419: gap of unknown length
 * 3420 4847: contig of 1428 bp in length
 * 4848 4947: gap of unknown length
 * 4948 8951: contig of 4004 bp in length
 * 8952 9051: gap of unknown length
 * 9052 13372: contig of 4321 bp in length
 * 13373 13472: gap of unknown length
 * 13473 28332: contig of 14860 bp in length
 * 28333 28432: gap of unknown length
 * 28433 64887: contig of 36455 bp in length
 * 64888 64987: gap of unknown length
 * 64988 158869: contig of 93882 bp in length
 * 158870 158969: gap of unknown length
 * 158970 321569: contig of 162600 bp in length
 * 321570 321669: gap of unknown length
 * 321670 322906: contig of 1237 bp in length.

FEATURES

source

1. 322906
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clones="RP24-66C10"
 1. 1483
 /note="assembly_name:Contig10"
 1584. 3319
 /note="assembly_name:Contig12"
 3420. 4847
 /note="assembly_name:Contig13"
 4948. 8951
 /note="assembly_name:Contig15"
 9052. 13372
 /note="assembly_name:Contig16"
 13473. 28332
 /note="assembly_name:Contig17"
 28433. 64887
 /note="assembly_name:Contig18"
 64988. 158869
 /note="assembly_name:Contig19"
 158970. 321569
 /note="assembly_name:Contig20"
 321670. 322906
 /note="assembly_name:Contig7"

misc_feature

1584. 3319
 /note="assembly_name:Contig10"
 3420. 4847
 /note="assembly_name:Contig12"
 4948. 8951
 /note="assembly_name:Contig13"
 9052. 13372
 /note="assembly_name:Contig15"
 13473. 28332
 /note="assembly_name:Contig16"
 28433. 64887
 /note="assembly_name:Contig17"
 64988. 158869
 /note="assembly_name:Contig18"
 158970. 321569
 /note="assembly_name:Contig19"
 321670. 322906
 /note="assembly_name:Contig20"

FEATURES

source

1. 161021
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"

ORIGIN

Query Match 8.5%; Score 47.6; DB 2; Length 322906;
 Best Local Similarity 61.1%; Pred. No. 0.35;
 Matches 77; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
 QY 418 AAGAACAGGAGACAGAGACAAATTCAAATGCTCTTACAGAGAGATGATTATATAGATA 477
 Db 23449 AATAGCAAGAAGCTGGAAAGAACCCAGATGTCCTCAACAGAGAGATGGATACAGAAAT 23390
 QY 478 ATGGTACATTCAATTAACATGATATATATTACTTAATTAATAAACATACCAACACACACAACA 537
 Db 23389 ATGGTTCATTACACATGGAGAACTACTCAGCTATTAAACCAATGAATCTACAAAATT 23330
 QY 538 CCTCGA 543
 Db 23329 CTTAGA 23324

RESULT 14

EX537253/c

LOCUS BX537253 161021 bp DNA linear ROD 04-OCT-2003
 DEFINITION Mouse DNA sequence from clone RP24-173B15 on chromosome 2, complete
 sequence.

ACCESSION BX537253

VERSION BX537253.6 GI:37518523

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 161021)
 Humphries, M.

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (02-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT

On Oct 5, 2003 this sequence version replaced gi:32812728.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to confirm this sequence. Sequence data
 from the whole genome shotgun alone has only been used where it has
 a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
 on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/Celegans/wormpep> RP24-173B15 is
 from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

Location/Qualifiers

1. 161021

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

```

: /chromosome="2"
: /clone="RP24-173B15"
: /clone_lib="RPC1-24"
:
ORIGIN
:
Query Match      8.4%  Score 47.4  DB 10  Length 161021;
Best Local Similarity 54.2%  Pred. No. 0.43;
Matches 96; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 356 ATGTATACATATTTAATACACAATGAGAAGCTCTTGGGGTGGTGGAGTGTATGGGC 415
DB 69393 ATATATATATATATATACAGAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGA 69334
QY 416 TGAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
DB 69333 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69274
QY 476 TAATGGTACATTCATATAAATGATATATTTACTTATTAATTAACATACCAACACACA 532
DB 69273 TACTGGGAGAAAATGAAACAAAACAAACAAACAAACAAACAAACAAACAAACAA 69217

RESULT 15
AC133842 220394 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-185E3, *** SEQUENCING IN PROGRESS ***
ACCESSION AC133842
VERSION AC133842.3 GI:30520933
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 220394)
AUTHORS Murny,D,Marie, Metker,M,Lee, Abramzon,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
Anylebechi,V, Ayodeji,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Cesar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Georgievski,B, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,
Gunnatre,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,B, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M,
Hollins,R, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpas,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,J, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorensheva,L, Louisegh,H, Lozado,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindartne,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Muridasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,S, Norris,S,
Nwakoileme,O, Okwono,G, Olarnpunsagoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkoch,C,
Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L-L,
Puazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,P,
Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFNZ
Center clone name: CH230-185E3
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 203828 bases at least Q40
Consensus quality: 208285 bases at least Q30
Consensus quality: 210214 bases at least Q20
Estimated insert size: 216328; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 220394: contig of 220394 bp in length.
* Location/Qualifiers
  1..220394
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-185E3"

```

```

misc_feature 1..3705
  /note="wgs_end_extension
  clone_end:T7"
  7885..8277
misc_feature
  /note="clone_boundary
  clone_end:T7
  site:EcoRI
misc_feature
  end_sequence:BH263108"
  complement(219195..219782)
  /note="clone_boundary
  clone_end:Sp6
  site:EcoRI
  end_sequence:BH263109"

ORIGIN
Query Match      8.3%; Score 47; DB 2; Length 220394;
Best Local Similarity 64.0%; Pred. No. 0.51;
Matches 71; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 418 AAGACAGGAGACAGACAAATTCAAATGCTTCCTTACACAGAGACTGATTATATAGATA 477
Db 45508 AATAGCCAGAAATTGTATACAAATTAGATGTTCTCTCAAGAGAGAGAAATAAATAAAT 45567
QY 478 ATGCTACATTTCATATAACATGATATATTTTACTAATTAATAACATACCAACA 528
Db 45568 ATGCTACATTTCATATAACATGATATATTTTACTAATTAATAACATACCAACA 528

```

Search completed: August 14, 2004, 19:01:42
 Job time : 2448.69 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 1929.83 Seconds
(without alignments)
8711.839 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563

Sequence: 1 ggtacatcttggtgtggat.....gocgctagctcagtagctag 563

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estlin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estci.*

9: gb_estli.*

10: gb_est2.*

11: gb_estc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.4	8.6	679	29	CE299099 tigr-gss-
C 2	48.2	8.6	535	10	BE635285 uv60a02.y
3	47.4	8.4	444	28	AZ003793 RPCI-23-3
4	46.6	8.3	459	28	AZ620342 LM0453u01

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	46	8.2	666	28	AQ933978
6	45.8	8.1	565	28	AQ933978
7	45.8	8.1	1201	9	AL553699
8	45.4	8.1	1101	29	CNS0106X
9	45	8.0	456	28	AQ556776
10	45	8.0	539	28	AQ556776
11	45	8.0	784	28	B2160642
12	45	8.0	794	13	EX097137
13	45	8.0	897	28	B2178903
14	45	8.0	3211	11	AK081359
15	44.4	7.9	345	28	BH069264
16	44.4	7.9	742	28	BZ092673
17	44.4	7.9	1201	9	AL548579
18	44.2	7.9	369	28	BH088169
19	44	7.8	395	28	AZ348426
20	43.8	7.8	1201	13	EX378989
21	43.4	7.7	184	28	AZ878343
22	43.4	7.7	370	9	AA96509
23	43.4	7.7	381	29	CE126378
24	43.4	7.7	388	28	AZ094991
25	43.4	7.7	464	28	AZ634360
26	43.4	7.7	492	28	AZ367086
27	43.4	7.7	499	9	AA926061
28	43.4	7.7	543	28	AZ873070
29	43.4	7.7	561	28	AZ278295
30	43.4	7.7	570	28	AZ284050
31	43.4	7.7	604	28	AZ375839
32	43.4	7.7	637	28	AQ923677
33	43.4	7.7	648	28	AZ376520
34	43.4	7.7	650	29	AG092150
35	43.4	7.7	796	28	AZ734214
36	43.4	7.7	897	28	BZ144279
37	43.2	7.7	514	28	AZ290946
38	43.2	7.7	535	28	AQ779549
39	43.2	7.7	574	28	BZ900086
40	43.2	7.7	600	28	AZ363255
41	43	7.6	782	28	BH042930
42	42.8	7.6	493	28	AZ431479
43	42.8	7.6	577	28	BH11528
44	42.8	7.6	617	28	AZ384488
45	42.8	7.6	629	28	AZ256692

ALIGNMENTS

RESULT 1
CE299099/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CE299099 tigr-gss-dog-1700035973606 Dog Library Canis familiaris genomic,
genomic survey sequence.

CE299099 CE299099.1 GI:36086776
GSS.

Canis familiaris (dog)
Canis familiaris
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

1 (bases 1 to 679)
Kirkness,E.F., Baina,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.

The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)

22875432
14512627

Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA

Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org

and selected for ampicillin resistance."

QY	4/8	ATGGTACATTCA	TATAA	AACATG	ATATAT	TTTACT	TAAT	TAAAA	510
Db	360	ATGGTACATT	TACACA	ATGA	AGTACT	ATTTCAG	CTATT	TAAAA	400

AQ333978	AQ333978	666 bp	DNA	linear	GSS 21-DEC-1999
LOCUS	RPCI-23-286P18.TV	Mus musculus genomic clone			
DEFINITION	RPCI-23-286P18	genomic survey sequence.			

REFERENCE	1. (Users 1, CO 000)
AUTHORS	Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other GSSES: RPCI-23-286P18.TJ

Department Of Zoology/Cellular Genomics
The Institute For Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RBC1-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering/frame.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac/ends/mouse/bac_end_intro.html
Plate: 286 row: P column: 18

BACPAC resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (<http://resgen.com>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 286 row: P column: 18
Seq primer: T7
Class: BAC ends.

FEATURES	Location/Qualifiers	source
1. .666		
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/strain="C57BL/6J"	
	/db_xref="taxon:10090"	
	/clone="RPC1-23-286F18"	
	/sex="Female"	
	/lab_host="DH10B"	
	/clone_lib="RPC1-33"	
	/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1: ECORI; Site 2: ECORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of ECORI and ECORI Methyase. Size selected DNA was cloned into the pBACE3.6 vector at the ECORI sites. The ligation products were transformed into DH10B electrocompetent cells. http://life-technologies.com	

ORIGIN

Query Match	8.2%	Score 46	DB 28	Length 666
Best Local Similarity	60.3%	Pred. No. 1.3		
Matches 76	Conservative	0	Mismatches 50	Indels 0
				Gaps 0

BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly); Genomic survey sequence.

AL098595
AL098595.1 Gi:5610206
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk>. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelosAcl1.

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBelosAcl1"
/note="end : T7"

ORIGIN
Query Match 8.1%; Score 45.4; DB 29; Length 1101;
Best Local Similarity 19.1%; Pred.No.1.;
Matches 99; Conservative 190; Mismatches 230; Indels 0; Gaps 0;

QY 26 TTGGACATACCTTTTATTTCCTTACCCTGTTCATCATCAAACTCTTAGGATGTATTACTTC 85
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
578 TKKHCTCCHKTKTTWTMTWKAKGTGTGTDKAKTKATWAGTGKTATRTWTAATKA 637
QY 86 CAAGGGCGTTAACTTATTCAGATTGACTGAGTCTCTTATTTTCCTTTAAATTTACTAGA 145
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
638 KRTADTKTRTATAATAKATKKKARKDKDTAKDAKWMDKATKAKAKKAAKAKAAAKT 697
QY 146 AGTGAGCTCCAGAACTACAGAAAATAGAGGAAGTCTCCAATTGAGGCATGAACGTGTG 205
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
698 ATGKDARAKAKAKATPKAKARAWDTATWTATADAADKGAKDKAKAKADAFDRK 757
QY 206 AGCACCTGGCATTTAAGCATGAAGAGTAGGGCTTCTATGGTAGGGAGCTGGATAGCAGC 265
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
758 RWDKDKRKRKAKAKAADADADAKADAKADAKADADADDGRGDGKKRKAKDKR 817
QY 266 ATTCCAGGAAGATCTCAGAGTCGAAACAATAGATTATCAGTTTAATCTCTTGGAC 325
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
818 KKXKDKAKGDKKKKAKDKKAAAKAKADADGAKAKERRAGDKDKAKADAKAAKAKAKDD 877
QY 326 CAAAGAGACCTTGAATCCTGGCTCGGTGATGTATACATATTTTATACAAATGAGAG 385
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
878 DDAAATKAKATKAKKDKAKAKKKKKKKDKAKAKAKAKADAKDKDDDKDDKDAK 937
QY 386 CTCCTTGGGGTGAGTGATAGGGCTGAAGAACAGGAGACAGAGAACAATTCAAA 445
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
938 DKKKGWKDSAKDKKDDDKDKDKDKDKDKADAKAWAKADADAKADAKAKAKAKAAD 997
QY 446 TGTCCTTTACAGAAAGACTGATTATATAGATATGCTATCATATCAATTAACATGATAT 505
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
998 DDAADAKADAKAKAKADDAKAKAKADAKADAKADAKADAKAAKAKAKADADA 1057
QY 506 TTACTTAATTAACATACCAACACACACACACCTCGAG 544
Db :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 10
AZ101750
LOCUS
DEFINITION
AZ101750 539 bp DNA linear GSS 09-MAY-2000
RPCI-23-452H21.TV RPCI-23 Mus musculus genomic clone
RPCI-23-452H21, genomic survey sequence.
ACCESSION
AZ101750
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS:
Mus musculus (house mouse)
REFERENCE
AUTHORS
1 (bases 1 to 539)
Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S.,
Akinret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de
Jong, P. and Fraser, C.M.
TITLE
JOURNAL
COMMENT
Mouse BAC End Sequences from Library RPCI-23
Other GSSs: RPCI-23-452H21.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 452 row: H column: 21
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..539
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-452H21"
/sex="Female"
/lab_host="PH108"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBAC3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN
Query Match 8.0%; Score 45; DB 28; Length 539;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 75; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 418 AAGACAGGACAGACAGCAATTCATTCCTTACACAGACACTGATTATATAGATA 477
DB 308 ATATGCCAGAGCTGGAAAGACCCAGATGTCCTTCAACAGAGGATGGATCAAAAAT 367
QY 478 ATGGTACATTCATATAACATGATATATTACTTAATTAACATCAACACACACACA 537
DB 368 GTGGTACATTCATATAACATGATATATTACTTAATTAACATCAACACACACA 427
QY 538 CCTCG 542
DB 428 CCTAG 432

RESULT 11

BZ160642

LOCUS
DEFINITION
BZ160642 784 bp DNA linear GSS 11-OCT-2002
CH230-327L4.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-327L4, genomic survey sequence.
ACCESSION
BZ160642
VERSION
KEYWORDS
SOURCE
ORGANISM
GSS:
Rattus norvegicus (Norway rat)
REFERENCE
AUTHORS
1 (bases 1 to 784)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, P., de Jong, P. and Fraser, C.M.
TITLE
JOURNAL
COMMENT
Rat BAC End Sequences from Library CHORI-230 MboI segment
Other GSSs: CH230-327L4.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/orering_information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 327 row: L column: 4
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..784
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-327L4"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pBAC3.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by
Pieter de Jong"
ORIGIN
Query Match 8.0%; Score 45; DB 28; Length 784;
Best Local Similarity 65.3%; Pred. No. 2.2;
Matches 66; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 418 AAGACAGGACAGACAGCAATTCATTCCTTACACAGACACTGATTATATAGATA 477
DB 499 AATACCCAGAAATGGAAACAATCGGATGTCCTTACACAGACACTGGATACATAAAT 558
QY 478 ATGGTACATTCATATAACATGATATATTACTTAATTAATAA 518
DB 559 GTGGTACATTCATATAACATGATATATTACTTAATTAATAA 599

RESULT 12
BX097137/c
LOCUS
DEFINITION
BX097137 794 bp mRNA linear EST 04-FEB-2003
IMAGD98804619_1 IMAGE:280131, mRNA sequence.
ACCESSION
BX097137
VERSION
KEYWORDS
SOURCE
ORGANISM
EST:
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
BX097137 Soares multiple sclerosis 2NDHMS Homo sapiens cDNA clone
IMAGD98804619_1 IMAGE:280131, mRNA sequence.

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, F., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076961
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6
 (Bases 1 to 3211)
 Fukuchi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Adachi, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishii, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, D., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp]
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. .3211
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:C130010D01"
 /db_xref="MGI:2413724"
 /db_xref="taxon:10090"
 /clone="C130010D01"
 /tissue_type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="16 days embryo"
 1. .3211
 misc_feature
 source
 FEATURES
 1. .345
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-257C9"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /clone_lib="RPCI-24"
 /notes="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 Query Match 7.9%; Score 44.4; DB 28; Length 345;

/note="hypothetical HMG-I and HMG-Y DNA-binding domain
 (A+T-hook) containing protein (InterPro|IPR000637,
 evidence: InterPro)"
 Query Match 8.0%; Score 45; DB 11; Length 3211;
 Best Local Similarity 60.0%; Pred. No. 1.8; Mismatches 0; Gaps 0;
 Matches 75; Conservative 0; Indels 50;

ORIGIN
 QY 418 AAGAACAGGAGACAGAGACAAATTCAAATGCTTACACAGACACATGATTATATACATA 477
 Db 2309 AATAGCCAGAAAGCTGGAAAGAACCCAGATGTCCTCAACAGAGATGAATATAAAAAAT 2368
 QY 478 ATGGTACATTCATATAACATGATATATTACTAATTAATAAACAACACACACACAACA 537
 Db 2369 ATGGTACATTTACACAAATGGTATATCTACTAGCAATTTAAAACAATGAATTCATGAATT 2428
 QY 538 CCTCG 542
 Db 2429 CTTAG 2433

RESULT 15
 BH069264
 LOCUS
 DEFINITION
 RPCI-24-257C9.TJ RPCI-24 Mus musculus genomic clone RPCI-24-257C9,
 genomic survey sequence.
 ACCESSION
 BH069264
 VERSION
 BH069264.1 GI:14888861
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 345)
 Zha, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
 Tsegaye, G., Geer, K., Krol, W., Shvartsbeyn, A., Gebregorgis, E.,
 Russell, D., de Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igr.org
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdjong@mail.cho.org). Clones may be purchased from BACPAC
 Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 257 row: C column: 9
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1. .345
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-257C9"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /clone_lib="RPCI-24"
 /notes="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBAC1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."
 Query Match 7.9%; Score 44.4; DB 28; Length 345;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 07:10:17 ; Search time 283.474 Seconds
(without alignments)
8437.229 Million cell updates/sec

Title: US-10-082-830-99
Perfect score: 563
Sequence: 1 ggtacatcttggtgtggtat.....gccgctagctcgagttag 563

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: Geneseq1380s.*
2: Geneseq1930s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002s.*
7: Geneseq2003as.*
8: Geneseq2003bs.*
9: Geneseq2003cs.*
10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	563	100.0	563	6	ABK93493 Human bre
2	43.4	7.7	75798	7	ABX77212 Mouse uri
3	42.4	7.5	445	4	AAI87875 Human pol
4	42.4	7.5	7374	4	AAI07519 Human rep
5	42.4	7.5	7374	4	ABA08211 Human ova
6	42	7.5	16766	6	ABL34157 Human imm
7	42	7.5	144792	9	ADC87620 Human GPC
8	42	7.5	349981	9	ADC87619 Human GPC
9	41.2	7.3	167343	6	ABL64403 Stomach c
10	41.2	7.3	167343	6	ABL67239 Thyroid c
11	40.8	7.2	77425	6	ABK83502 Human cDN
12	40.2	7.1	17032	6	AAB84885 Human imm
13	40.2	7.1	173810	6	ABN85752 Mouse chr
14	40	7.1	11027	4	AAI37290 Human mus
15	40	7.1	11027	4	AAI37289 Human mus
16	40	7.1	11027	7	ABX60277 cDNA enco
17	40	7.1	11027	7	ABX60278 cDNA enco
18	40	7.1	104000	8	AAI57669 Human pho
19	39.6	7.0	21837	4	AAK85946 Human imm
20	39.4	7.0	323	4	AAI36295 Human mus
21	39.4	7.0	323	4	AAI36293 Human mus
22	39.4	7.0	323	7	ABX59283 cDNA enco
23	39.4	7.0	323	7	ABX59281 cDNA enco

C	24	39.4	7.0	1771	5	AAH22966 Human pho
	25	39.4	7.0	2890	4	AAH77031 Human NF-
	26	39.4	7.0	110000	5	Aaf84800 Nucleotid
C	27	39.2	7.0	9725	6	ABL33292 Human imm
C	28	39.2	7.0	9725	6	ABL33292 Human che
	29	39.2	7.0	21537	6	ABL33999 Human imm
C	30	39	6.9	606	4	AAK63763 Human imm
C	31	39	6.9	4527	4	AAK85875 Human imm
C	32	39	6.9	81001	4	AAI30035 Human apo
C	33	38.8	6.9	179	2	AAI19669 Human gen
C	34	38.8	6.9	512	4	AAI86036 Human pol
C	35	38.8	6.9	9507	7	AAI07097 Human rep
C	36	38.8	6.9	9507	7	ABZ57483 Human sec
C	37	38.8	6.9	9507	7	ABZ73900 Secreted
	38	38.6	6.9	768	4	AAK86425 Human imm
	39	38.6	6.9	768	4	AAK86423 Human imm
	40	38.6	6.9	768	4	AAK86426 Human imm
	41	38.6	6.9	16766	6	ABL34156 Human imm
C	42	38.6	6.9	173810	6	ABN85752 Mouse chr
C	43	38.4	6.8	14919	4	AAI46506 Human su
	44	38.4	6.8	33923	4	AAK67071 Human imm
	45	38.4	6.8	149671	6	ABK84797 Human cDN

ALIGNMENTS

RESULT 1
ABK93493
ID ABK93493 standard; cDNA; 563 BP.
XX
AC ABK93493;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human breast specific nucleic acid, BSNA #99.
XX
KW Human; ss; breast specific nucleic acid; BSNA; breast cancer;
XX mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.
XX
OS Homo sapiens.
XX
PN WO200236807-A2.
XX
PD 10-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-US046888.
XX
PR 27-OCT-2000; 2000US-0243802P.
XX
PS (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Salceda S, Liu C, Turner LR;
XX
PI WPI; 2002-463415/49.
XX
PT New breast-specific nucleic acids and polypeptides, useful for
PT identifying, diagnosing, monitoring, staging, imaging, and treating
PT breast cancer and non-cancerous disease states in breast tissues.
XX
PS Claim 1; Page 191-192; 281pp; English.
XX
CC The invention relates to breast specific nucleic acids (BSNA) and breast-
CC specific polypeptides (BSP). Also included are a method for determining
CC the BSNA in a sample, a vector comprising a BSNA, a host cell comprising
CC the vector, a method for producing a polypeptide encoded by a BSNA, an
CC anti-BSP antibody and a method for determining the presence of a BSP in a
CC sample. The breast-specific nucleic acids, polypeptides and compositions
CC comprising them are useful for identifying, diagnosing, monitoring,
CC staging, imaging, and treating breast cancer, mammary tumour and non-
CC cancerous disease states in breast tissue; for identifying breast tissue;
CC for monitoring, identifying and/or designing agonists and antagonists of
CC the polypeptides; in gene therapy; in producing transgenic animals and


```

OS Homo sapiens.
XX WO200164835-A2.
XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US004927.
XX 28-FEB-2000; 2000US-00515126.
XX 18-MAY-2000; 2000US-00577409.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Dmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO07944.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX Claim 1; SEQ ID NO 7935; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 445 BP; 158 A; 75 C; 109 G; 103 T; 0 U; 0 Other;
XX
Query Match 7.5%; Score 42.4; DB 4; Length 445;
Best Local Similarity 62.0%; Pred. No. 0.22;
Matches 67; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
Qy 430 CAGAGACAAATTCAAATGTCCTTACACAGAGACTGATTATATAGATAATGGTACATTCA 489
Db 148 CTGGAACAACCTCAATGTCCTTAAACAGGAGATGATGAAGTTAATGGTATTCA 207
Qy 490 TATAACATGATATATTACTAATTAACATACCAACACACACACA 537
Db 208 CATAATAGAACTACTTGTGACATAAAGAGATAAAGTCCTGGAACA 255
RESULT 4
AAL07519/c
ID AAL07519 standard; DNA; 7374 BP.
XX
XX AAL07519;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 10207.
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0011339.
XX
31-JAN-2000; 2000US-0179065P.
04-FEB-2000; 2000US-0180628P.
24-FEB-2000; 2000US-0184664P.
02-MAR-2000; 2000US-0186350P.
16-MAR-2000; 2000US-0189874P.
17-MAR-2000; 2000US-0190076P.
18-APR-2000; 2000US-0198123P.
19-MAY-2000; 2000US-0205515P.
07-JUN-2000; 2000US-0209467P.
28-JUN-2000; 2000US-0214886P.
30-JUN-2000; 2000US-0215135P.
07-JUL-2000; 2000US-0216647P.
07-JUL-2000; 2000US-0216880P.
11-JUL-2000; 2000US-0217487P.
14-JUL-2000; 2000US-0218290P.
26-JUL-2000; 2000US-0220963P.
26-JUL-2000; 2000US-0220964P.
14-AUG-2000; 2000US-0224518P.
14-AUG-2000; 2000US-0224519P.
14-AUG-2000; 2000US-0225213P.
14-AUG-2000; 2000US-0225214P.
14-AUG-2000; 2000US-0225266P.
14-AUG-2000; 2000US-0225267P.
14-AUG-2000; 2000US-0225268P.
14-AUG-2000; 2000US-0225270P.
14-AUG-2000; 2000US-0225447P.
14-AUG-2000; 2000US-0225757P.
14-AUG-2000; 2000US-0227009P.
30-AUG-2000; 2000US-0228924P.
01-SEP-2000; 2000US-0229287P.
01-SEP-2000; 2000US-0229343P.
01-SEP-2000; 2000US-0229344P.
01-SEP-2000; 2000US-0229345P.
05-SEP-2000; 2000US-0229509P.
05-SEP-2000; 2000US-0229513P.
06-SEP-2000; 2000US-0230437P.
06-SEP-2000; 2000US-0230438P.
08-SEP-2000; 2000US-0231243P.
08-SEP-2000; 2000US-0231244P.
08-SEP-2000; 2000US-0231413P.
08-SEP-2000; 2000US-0231414P.
08-SEP-2000; 2000US-0232080P.
08-SEP-2000; 2000US-0232081P.
12-SEP-2000; 2000US-0231968P.
14-SEP-2000; 2000US-0232397P.
14-SEP-2000; 2000US-0232398P.
14-SEP-2000; 2000US-0232399P.
14-SEP-2000; 2000US-0232400P.
14-SEP-2000; 2000US-0232401P.
14-SEP-2000; 2000US-0233063P.
14-SEP-2000; 2000US-0233064P.
14-SEP-2000; 2000US-0233065P.
21-SEP-2000; 2000US-0234223P.
21-SEP-2000; 2000US-0234274P.
25-SEP-2000; 2000US-0234997P.
25-SEP-2000; 2000US-0234998P.
26-SEP-2000; 2000US-0235484P.
27-SEP-2000; 2000US-0235834P.
27-SEP-2000; 2000US-0235836P.
29-SEP-2000; 2000US-0236327P.
29-SEP-2000; 2000US-0236372P.
29-SEP-2000; 2000US-0236376P.
29-SEP-2000; 2000US-0236368P.
29-SEP-2000; 2000US-0236369P.
29-SEP-2000; 2000US-0236370P.
02-OCT-2000; 2000US-0236802P.

```

PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259567P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX

PS Disclosure; SEQ ID NO 10207; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 7374 BP; 1974 A; 1512 C; 1461 G; 2427 T; 0 U; 0 Other;

Query Match 7.5%; Score 42.4; DB 4; Length 7374;
Best Local Similarity 57.6%; Pred. No. 0.53;
Matches 76; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 387 TCCTGAGGGTGAGTGAGTGAGGCTGCAAGCAACAGGAGACAGAGCAATTCAAAT 446
Db 1091 TACATGAATGTTGTGAGCAATATACATTAATGACCAAAAATGAATAATCAAAT 1032
QY 447 GTCTTACACAGAAAGACTGATTATATAGATAATCGTACATTTCATATAAATGATATATT 506
Db 1031 GTCCATCAACTGATGATAGATAATAAATGATCATATATCCATACATGAAATGTTATT 972
QY 507 TACTAATTAATAA 518
Db 971 CAGCAATAAAAA 960

RESULT 5
ABA08211/c
ID ABA08211 standard; DNA; 7374 BP.
XX
AC ABA08211;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human ovarian and breast cancer associated polynucleotide SEQ ID NO 1006.
XX
KW Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiogenic; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ds.
XX
OS Homo sapiens.
XX
PN WO200155325-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001345.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217498P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.

Query Match 7.5%; Score 42.4; DB 4; Length 7374;
 Best Local Similarity 57.6%; Pred. No. 0.53;
 Matches 76; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 387 TCCTTGAGGTGAGTGTAGGCTGGAAGACAGAGACAGACAAATTCAAAT 446
 Db 1091 TACATGAATGTTGTTGAAGCATTATACATAATGACCAAAAAATGAAATTAATTCAAAT 1032

QY 447 GTCCTTACAGAGAGCTGATTATATAGATAAATGGTACATTCATATAAACAATGATATATT 506
 Db 1031 GTCCTTACAGAGAGCTGATTATATAGATAAATGGTACATTCATATAAACAATGATATATT 506

QY 507 TACTAATTAATAA 518
 Db 971 CAGCAATAAAAA 960

RESULT 6
 ABL34157/c
 ID ABL34157 standard; DNA; 16766 BP.
 XX AC ABL34157;
 XX DT 26-MAR-2002 (first entry)
 XX DE Human immune system associated gene SEQ ID NO: 2130.
 XX DX Human; immune system disease; cytosine methylation; antiasthmatic;
 XX KW antiarteriosclerotic; antianemic; cytostatic; nootropic;
 XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 XX KW antirheumatic; antiarthritic; anidiabetic; antipsoriatic;
 XX KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 XX KW ds.
 XX OS Homo sapiens.
 XX PN WO200200928-A2.
 XX PD 03-JAN-2002.
 XX PF 02-JUL-2001; 2001WO-EP007537.
 XX PR 30-JUN-2000; 2000DE-01032529.
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX Nucleic acid' comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX Claim 1; SEQ ID NO 2130; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX Sequence 16766 BP; 4604 A; 267 C; 3632 G; 8263 T; 0 U; 0 Other;

Query Match 7.5%; Score 42; DB 6; Length 16766;
 Best Local Similarity 60.5%; Pred. No. 0.88;

Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 422 ACAGGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATTATAGTATG 481
 Db 503 AAAAAAATAAATAAATTCAAATATCTATAATAAAAAATCTAATTAATAATAA 444

QY 482 TACATTCATATAAACAATGATATATTTACTAATTAATAAACAATCAACACACAA 535
 Db 443 TATATCTATAAATAAACAATGATATATTTACTAATTAATAAACAATCAACAA 390

RESULT 7
 ADC87620/c
 ID ADC87620 standard; DNA; 144792 BP.
 XX AC ADC87620;
 XX DT 01-JAN-2004 (first entry)
 XX DE Human GPCR related polynucleotide SEQ ID NO:2073.
 XX DX ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 XX KW gene therapy.
 XX OS Homo sapiens.
 XX PN EP1270724-A2.
 XX PD 02-JAN-2003.
 XX PF 18-JUN-2002; 2002EP-00013517.
 XX PR 18-JUN-2001; 2001JP-00246789.
 XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX PI Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX WPI; 2003-315783/31.
 XX New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of, increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX Disclosure; SEQ ID NO 2073; 28pp; English.
 XX The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC87618-ADC87623 represent polynucleotide sequences
 CC related to the invention.
 XX Sequence 144792 BP; 39827 A; 32142 C; 33413 G; 39310 T; 0 U; 100 Other;

Query Match 7.5%; Score 42; DB 9; Length 144792;
 Best Local Similarity 66.7%; Pred. No. 1.7;
 Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 426 GACACAGACAGAAATTCAAATGTCCTTACACAGAACTGATTATAGATAATG 485
 Db 9727 GAAAGAGAGAGCAATCTAGATGTCCTCAATAGGAGACTGTTGATATAAATCTGCT 9668

QY 486 TTCATATAAACAATGATATATTTACTAATTA 515
 Db 9667 TGCATCTATTCATGAGATATTTAGATTTA 9638

RESULT 8
 ADC87619/c

CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1 and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 XX
 SQ Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 0 U; 13 Other;
 Query Match 7.3%; Score 41.2; DB 6; Length 167343;
 Best Local Similarity 67.4%; Pred. No. 3;
 Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
 QY 409 ATAGGGCTGAAGAACAGGAGACAGACAAATTCATTCCTTACACAGAGACTGATT 468
 Db 159637 ATTTTCAAAATACAAAGATTGGAAGAGGCAATATCTTGGTAGTAGACTGATG 159696
 QY 469 ATATAGATTAATGGTACATTCATATA 494
 Db 159697 AAATACATTGTGCTACATACATACAA 159722
 RESULT 10
 ABL67239
 ID ABL67239 standard; DNA; 167343 BP.
 XX
 AC ABL67239;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Thyroid cancer related gene sequence SEQ ID NO:5576.
 XX
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO200194629-A2.
 XX
 FD 13-DEC-2001..
 XX
 PF 30-MAY-2001; 2001WO-US010838.
 XX
 PR 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-023313P.
 PR 18-SEP-2000; 2000US-023361P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 20-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.

PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 02-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237588P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 PA (AVAL-) AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.
 XX
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 PS Claim 1; SEQ ID NO 5576; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
 CC tumour
 XX
 SQ Sequence 167343 BP; 49550 A; 37098 C; 35753 G; 44929 T; 0 U; 13 Other;
 Query Match 7.3%; Score 41.2; DB 6; Length 167343;
 Best Local Similarity 67.4%; Pred. No. 3;
 Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
 QY 409 ATAGGGCTGAAGAACAGGAGACAGACAAATTCATTCCTTACACAGAGACTGATT 468
 Db 159637 ATTTTCAAAATACAAAGATTGGAAGAGGCAATATCTTGGTAGTAGACTGATG 159696
 QY 469 ATATAGATTAATGGTACATTCATATA 494
 Db 159697 AAATACATTGTGCTACATACATACAA 159722
 RESULT 11
 ABL67239/c
 ID ABL67239 standard; cDNA; 77425 BP.
 XX
 AC ABL67239;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #73.

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US030821.
XX
XX 03-OCT-2000; 2000US-0237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
XX Genes associated with granulocyte activation, which serves as diagnostic
XX markers that is useful for monitoring disease states and drug toxicity.
XX
XX Claim 1; SEQ ID NO 73; 11app; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing the
XX expression level to an expression level in an unactivated GC, where
XX differential expression of Gs is indicative of GCA. Also included are
XX modulating (M2) GA by contacting GC with an agent that alters the
XX expression of at least one gene in Gs; (2) screening (M3) for an agent
XX capable of modulating GCA or an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease using the gene expression
XX profile; (3) detecting (M4) an inflammation (especially chronic) in a
XX tissue, an allergic response in a subject, exposure of a subject to a
XX pathogen or sterile inflammatory disease, by detecting the level of
XX expression in a sample of the tissue of gene(s) from Gs, where the level
XX of expression of the gene is indicative of inflammation; (4) treating
XX (M5) an inflammation (especially chronic) or in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease, by contacting a tissue having inflammation with an
XX agent that modulates the expression of gene(s) from Gs in the tissue. M1
XX is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful
XX for screening an agent capable of modulating GCA preferably in an
XX inflammation in a tissue; M4 is useful for detecting an inflammation
XX (especially chronic) in a tissue, an allergic response in a subject,
XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, adult
XX cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
XX respiratory distress syndrome, inflammatory bowel disease, Crohn's
XX disease, ulcerative colitis, periodontal disease; also bacterial
XX infection, viral infection, parasitic infection, protozoal infection,
XX fungal infection and M5 is useful for treating one of the above
XX conditions. The present sequence represents a gene differentially
XX expressed in granulocytes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 77425 BP; 23611 A; 13793 C; 14665 G; 25356 T; 0 U; 0 Other;
Query Match 7.2%; Score 40.8; DB 6; Length 77425;
Best Local Similarity 59.5%; Pred. No. 3;
Matches 69; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 417 GAAGAAACAGGAGACAGAGCAATTCAAATGTCCTTACACAGAGACTGATTATATAGAT 476
Db 39110 GCATAGCCTCAAACTAGAAACAATCCAAATGTCCTTACAGAGAGCTAACAAATAAT 39051
QY 477 AATGGTACATTCATATAAACAATGATATATTTACTATTTAAACATACCAACACACA 532
Db 39050 GGTGGTATATTCATATATATACCTTATATCTATGGAATATAACAATGAGAATGCATA 38995
RESULT 12
AAK84885
ID AAK84885 standard; DNA; 17032 BP.
XX
XX AAK84885;
AC
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39697.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 14-JUL-2000; 2000US-0217496P.
XX 26-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225575P.
XX 14-AUG-2000; 2000US-0225758P.
XX 18-AUG-2000; 2000US-0225759P.
XX 22-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226688P.
XX 23-AUG-2000; 2000US-0227182P.
XX 30-AUG-2000; 2000US-0227009P.
XX 01-SEP-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 05-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.

DE Mouse chromosome 11 BAC clone RP23-270L8 SEQ ID NO 10.
XX Mouse; Can 1; antiinfertility; gynaecological; infertility;
KW premature ovarian failure; menopause; Sertoli Cell only syndrome;
KW BAC clone RP23-270L8; GenBank reference AC083815; chromosome 11; ds.
XX
OS Mus musculus.
XX
PN US2002119929-A1.
XX
PD 29-AUG-2002.
XX
PF 02-NOV-2001; 2001US-00003806.
XX
PP 03-NOV-2000; 2000US-0245872P.
XX
PR (BISH/) BISHOP C E.
PA (AGOU/) AGOULNIK A I.
PA (ZHUQ/) ZHU Q.
XX
PI Bishop CE, Agoulnik AI, Zhu Q;
XX
DR WPI; 2002-618953/66.
XX
PT A nucleic acid molecule (I) encoding a Can 1 polypeptide used in treating
PT infertility.
XX
XX Disclosure; Page; 45pp; English.
XX
XX The invention relates to a nucleic acid molecule (I) encoding a Can 1
CC polypeptide. The Can 1 nucleic acid molecule is used to diagnose or treat
CC infertility or premature ovarian failure or Sertoli Cell only syndrome in
CC a mammal. The present sequence is that of a mouse Can 1 encoding
CC chromosome 11 BAC clone RP23-270L8 of the invention. Note: The present
CC sequence is not given in the printed specification but was obtained
CC through the GenBank reference AC083815
XX
SQ Sequence 173810 BP; 50474 A; 33519 C; 33994 G; 55823 T; 0 U; 0 Other;
Query Match 7.1%; Score 40.2; DB 6; Length 173810;
Best Local Similarity 57.6%; Pred. No. 5.6;
Matches 72; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 418 AAGAACAGGAGACAGACAAATTCAAATGTCCTTACACAGAGACTGATTATATAGATA 477
Db 42876 AATAGCCAGACCTGGAAACACCAGATGCTTTTAAACAGGATAGATACAGAAAT 42935
Qy 478 ATGGTACATTCATTAACATGATATATTTACTAATTAACACATACCACACACACAACA 537
Db 42936 ATGGTACATTTACAAATGGAGTACTATTTCAGCTATTTAAAGATGCAATTCATGAAAT 42995
Qy 538 CCTCG 542
Db 42996 CTTAG 43000
RESULT 14
AAL37290
ID AAL37290 standard; DNA; 11027 BP.
XX
AC AAL37290;
XX
DT 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polynucleotide SEQ ID NO 3655.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW candiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX

OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001338.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189974P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216980P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226689P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.

XX	PI	Rosen CA, Barash SC, Ruben SM;
XX	WPI	2001-451937/48.
XX	PT	Isolated polypeptide for treating, preventing and/ or prognosing disorders related to the musculoskeletal system including musculoskeletal cancers and also for testing and detection e.g. diagnosis.
XX	PS	Example 2; SEQ ID NO 3655; 781pp + Sequence Listing; English.
XX	CC	The invention relates to novel genes (AAL34669-AAL37666) and proteins (AB03087-AB04109) associated with the musculoskeletal system useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	SQ	Sequence 11027 BP; 2851 A; 2179 C; 2539 G; 3458 T; 0 U; 0 Other;
Query Match 7.1%; Score 40; DB 4; Length 11027;		
Best Local Similarity 55.9%; Pred. No. 2.7;		
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;		
QY	417	GAAGAACAGGACAGACACAAATTCATTCCTTACACAGACAGCTGATTATAGAT 476
DB	5908	GAGGAACAAAGCTTAAAGACATCAAAATGCTCTTAAATAGAGACTGTCTTAATCAAT 5967
QY	477	AATGTCATTCATATAACATGATATATTTACTAATTAACATACCAACACACACAAC 536
DB	5968	TAGGCGATTCATTCAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 6027
QY	537	ACCTCGAGCGCTAGT 552
DB	6028	TGAGGAGGCTTTCAGT 6043
RESULT 15		
AAL37289		
ID	AAL37289	standard; DNA; 11027 BP.
XX	AAL37289;	
XX	08-JAN-2002	(first entry)
XX	Human musculoskeletal system related polynucleotide SEQ ID NO 3654.	
XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
XX	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antifungal;	
XX	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;	
XX	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;	
XX	neurological disease; infection; human; secreted protein;	
XX	musculoskeletal system; ds.	
XX	Homo sapiens.	
XX	WO200155367-A1.	
XX	02-AUG-2001.	
XX	17-JAN-2001; 2001WO-US001336.	

XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0196350P.
PR 16-MAR-2000; 2000US-0198874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-022547P.
PR 14-AUG-2000; 2000US-022557P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 22-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227709P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-023423P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 26-SEP-2000; 2000US-0234988P.
PR 27-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.

PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-024527P.
PR 08-NOV-2000; 2000US-024527P.
PR 08-NOV-2000; 2000US-024528P.
PR 08-NOV-2000; 2000US-024532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0248207P.
PR 17-NOV-2000; 2000US-0248208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
XX (HJMA-) HUNAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-451937/48.

PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the musculoskeletal system including musculoskeletal

PT cancers and also for testing and detection e.g. diagnosis.
XX
PS Example 2; SEQ ID NO 3654; 781pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful for
CC preventing, treating or ameliorating medical conditions e.g. by protein
CC or gene therapy. The genes are isolated from a range of human tissues
CC disclosed in the specification. The nucleic acids, proteins, antibodies
CC and (ant)agonists are useful in the diagnosis, treatment and prevention
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11027.BP; 2851 A; 2182 C; 2537 G; 3457 T; 0 U; 0 Other;
Query Match 7.1%; Score 40; DB 4; Length 11027;
Best Local Similarity 55.9%; Pred. No. 2.7;
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 417 GAAGAACAGGAGACAGAGACAAATCAAAATGCTCTTACAGAGAGACTGATTATATAGAT 476
Db |||||
QY 477 AATGGTACATTATATATAACATGATATATTACTAATTAATAACATACCACACACAC 536
Db |||||
QY 537 ACCTCGAGCGCGTAGT 552
Db |||||
QY 6028 TGAGGGAGCTTTCAGT 6043
Db |||||

Search completed: August 14, 2004, 19:13:46
Job time : 287.474 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 19:13:54 ; Search time 319.797 Seconds
(without alignments)
8638.053 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563

Sequence: 1 GGCACATCTGGCTGGAT.....GCCGTAGTCTGAGTCTAG 563

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3225727 seqs, 2453303834 residues

Total number of hits satisfying chosen parameters: 5451454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09D_PUBCOMB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09E_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	563	100.0	563	15	US-10-082-830-99
2	511.4	90.8	330973	13	Sequence 99, Appl
3	42.4	7.5	7374	10	Sequence 1498, Ap
4	42.4	7.5	7374	15	Sequence 10207, A
5	42	7.5	16766	15	Sequence 1006, Ap
6	42	7.5	744802	16	Sequence 2130, Ap
7	41.8	7.4	915	13	Sequence 1369, Ap
8	41.8	7.4	915	16	Sequence 120056,
9	41.8	7.4	154817	16	Sequence 120056,
10	41.2	7.3	167343	9	Sequence 334, App
11	41.2	7.3	167343	9	Sequence 281, App
12	41	7.3	97415	17	Sequence 273, App
13	40.2	7.1	414	13	Sequence 287, App
14	40.2	7.1	414	16	Sequence 16076, A
					Sequence 16076, A

15	40.2	7.1	509	13	US-10-240-425-622	Sequence 622, App
16	40.2	7.1	145068	17	US-10-322-281-33	Sequence 33, Appl
17	40.2	7.1	155579	16	US-10-085-117-283	Sequence 283, App
18	40.2	7.1	173808	14	US-10-003-806-10	Sequence 10, Appl
19	40	7.1	714	13	US-10-027-632-114034	Sequence 114034,
20	40	7.1	714	13	US-10-027-632-114035	Sequence 114035,
21	40	7.1	714	13	US-10-027-632-114036	Sequence 114036,
22	40	7.1	714	16	US-10-027-632-114034	Sequence 114034,
23	40	7.1	714	16	US-10-027-632-114035	Sequence 114035,
24	40	7.1	714	16	US-10-027-632-114036	Sequence 114036,
25	40	7.1	11027	9	US-09-764-877-3654	Sequence 3654, Ap
26	40	7.1	11027	9	US-09-764-877-3655	Sequence 3655, Ap
27	40	7.1	11027	16	US-10-242-515-3654	Sequence 3654, Ap
28	40	7.1	11027	16	US-10-242-515-3655	Sequence 3655, Ap
29	40	7.1	104000	15	US-10-012-984-14	Sequence 14, Appl
30	40	7.1	104000	17	US-10-673-523-14	Sequence 14, Appl
31	39.8	7.1	878	13	US-10-027-632-31606	Sequence 31606, A
32	39.8	7.1	878	16	US-10-027-632-31606	Sequence 31606, A
33	39.6	7.0	911	13	US-09-854-867-341	Sequence 341, App
34	39.4	7.0	323	9	US-09-764-877-2658	Sequence 2658, Ap
35	39.4	7.0	323	9	US-09-764-877-2660	Sequence 2660, Ap
36	39.4	7.0	323	16	US-10-242-515-2658	Sequence 2658, Ap
37	39.4	7.0	323	16	US-10-242-515-2660	Sequence 2660, Ap
38	39.4	7.0	1771	15	US-10-181-590-18	Sequence 18, Appl
39	39.2	7.0	9725	15	US-10-311-455-1265	Sequence 1265, Ap
40	39.2	7.0	21537	15	US-10-311-455-1972	Sequence 1972, Ap
41	39	6.9	870	13	US-10-027-632-160076	Sequence 160076,
42	39	6.9	870	13	US-10-027-632-160077	Sequence 160077,
43	39	6.9	870	13	US-10-027-632-160078	Sequence 160078,
44	39	6.9	870	16	US-10-027-632-160076	Sequence 160076,
45	39	6.9	870	16	US-10-027-632-160077	Sequence 160077,

ALIGNMENTS

RESULT 1

US-10-082-830-99
; Sequence 99, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 99
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-082-830-99

Query Match 100.0%; Score 563; DB 15; Length 563;
Best Local Similarity 100.0%; Pred. No. 4.1e-147;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCATCTTGCTGGATGGAAATTTGACATCTTTTCTTTCTTCTGTTACAT 60
Db 1 GGTCATCTTGCTGGATGGAAATTTGACATCTTTTCTTTCTTCTGTTACAT 60
QY 61 ATCAATCTTAGATGTATTACTTCCAGGGCGGTAAATCTTATTCAGATGACTGATC 120
Db 61 ATCAATCTTAGATGTATTACTTCCAGGGCGGTAAATCTTATTCAGATGACTGATC 120

Qy	121	TCTATTTCCTTAATTTTACATGAAGTGGAGCTCCCAAGAACTACAGAAATAGAGGAA	180
Db	121	TCTATTTCCTTAATTTTACTAGAAGTGGAGCTCCCAAGAACTACAGAAATAGAGGAA	180
Qy	181	AGTCTCCATTGAGCCATGAACGTGTAGCACCTGGCATTTAAGCATGAAGTAGTAGGCGTTC	240
Db	181	AGTCTCCATTGAGCCATGAACGTGTAGCACCTGGCATTTAAGCATGAAGTAGTAGGCGTTC	240
Qy	241	TATGTTAGGAGCTGGAGTAGGCAGCATTTCCAGGAAAGGATCTCAGAGGTCAGAAAACAATA	300
Db	241	TATGTTAGGAGCTGGAGTAGGCAGCATTTCCAGGAAAGGATCTCAGAGGTCAGAAAACAATA	300
Qy	301	GATTATCAGTTAAATACCTCTGGACCAAGAGACCTTGAAATCCTGGCTCGGTGATGTA	360
Db	301	GATTATCAGTTAAATACCTCTGGACCAAGAGACCTTGAAATCCTGGCTCGGTGATGTA	360
Qy	361	TACATATTTAATACAAATGAGAACTCCTTGAGGCTGAGTGAGAGTGATAGGGCTGAAG	420
Db	361	TACATATTTAATACAAATGAGAACTCCTTGAGGCTGAGTGAGAGTGATAGGGCTGAAG	420
Qy	421	AACAGAGACAGAGACAAATTCAAATGTCCTTACACAGAGACGTGATTATATAGATAATG	480
Db	421	AACAGAGACAGAGACAAATTCAAATGTCCTTACACAGAGACGTGATTATATAGATAATG	480
Qy	481	GTACATTTCATATAAATCATGATATATTTACTAATTTAAAAACATATCCCAACACACACACCT	540
Db	481	GTACATTTCATATAAATCATGATATATTTACTAATTTAAAAACATATCCCAACACACACACCT	540
Qy	541	CGAGCGCTAGTCTCGAGTCTAG	563
Db	541	CGAGCGCTAGTCTCGAGTCTAG	563

RESULT 2

```

US0001.12
; Sequence 1498, Application US/10087192
; Publication No. US20030182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1498
; LENGTH: 330973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(330973)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1498

```

	Query Match	90.8%;	Score 511.4;	DB 13;	Length 330973;
	Best Local Similarity	99.6%;	Pred. No. 4.2e-131;		
	Matches 523;	Conservative	0;	Mismatches 1;	Indels 1; Gaps 1;
QY	1	GGTACATCTTGGCTCTGGATGGAAATTTGACATCTTTTATTTTCTTTTACCTGTTACAT	60		
Db	147784	GGTACATCTTGGCTCTGGATGGAAATTTGACATCTTTTATTTTCTTTACCTGTTACAT	147843		
QY	61	ATCAAAATCTTAGGATGTATTCTTCCAAAGCGGGTTAAACTTATTCAAGATTGACTGAGTC	120		
Db	147844	ATCAAAATCTTAGGATGTATTCTTCCAAAGCGGGTTAAACTTATTCAAGATTGACTGAGTC	147903		

Qy	121	TCTATATTTTCTTAAATTTACTAGAACTGGAGCTCCGAGAACTACACAGAAATAGAGGAA	180
Db	147904	TCTATATTTTCTTAAATTTACTAGAACTGGAGCTCCGAGAACTACACAGAAATAGAGGAA	147963
Qy	181	AGTCTCCATTGAGCCATGAACCTGTGAGCACCTGGCATTTAAGCATGAAGAGTAGGCGCTTC	240
Db	147964	AGTCTCCATTGAGCCATGAACCTGTGAGCACCTGGCATTTAAGCATGAAGAGTAGGCGCTTC	148023
Qy	241	TATGTTAGGAGCTGGAGTAGGACAGCATTCACAGGAAGGATCTCAGAGGTCAGAAACAATA	300
Db	148024	TATGTTAGGAGCTGGAGTAGGACAGCATTCACAGGAAGGATCTCAGAGGTCAGAAACAATA	148083
Qy	301	GATTATCAGTTAAATATCTCTTGGACCAAGAGACCTTTGAAATCCTTGGCTCGGTGATGTA	360
Db	148084	GATTATCAGTTAAATATCTCTTGGACCAAGAGACCTTTGAAATCCTTGGCTCGGTGATGTA	148143
Qy	361	TACATATTTAATACACAATGAGAGGCTCCTTGAGGGGTGAGTGAGAGTGATAGGGCTGAAG	420
Db	148144	TACATATTTAATACACAATGAGAGGCTCCTTGAGGGGTGAGTGAGAGTGATAGGGCTGAAG	148203
Qy	421	RACAGGACACAGACAGCAATTCRAATCTCCTTACACAGAGAGCTGATTATATAGATAATG	480
Db	148204	RACAGGACACAGACAGCAATTCRAATCTCCTTACACAGAGAGCTGATTATATAGATAATG	148263
Qy	481	GTACATTTCCATATAAACATGATATATTTACTAATTTAAACATACCA	525
Db	148264	GTACATTTCCATAT - AACATGATATATTTTACTAATTTAAACATACCA	148307

RESULT 3

US-09-764-891-10207/c

; Sequence 10207, Application US/09764891

; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10207

; LENGTH: 7374

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-891-10207

RESULT 3

```

US-09-764-891-10207/C
; Sequence 10207, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10207
; LENGTH: 7374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-10207

```

	Query Match	7.5%;	Score 42.4;	DB 10;	Length 7374;
	Best Local Similarity	57.6%;	Pred. No. 1.1;	56;	Indels 0;
	Matches 76;	Conservative 0;	Mismatches		Gaps 0;
QY	387	TCCTTGGGGGTGAGTGATGATGGCTGGAAGAACGAGACAGACAGCAATTCAAAT	446		
DB	1091	TACATGAATGTTCTTGAAGCATTATACATAATGACCAAAAAAATGAAATTAATTCAAAT	1032		
QY	447	GFCCCTTACACAGAGAGCTGATTATATAGATAAATGGTGACATTCATATAAACATGATATATT	506		
DB	1031	GTTCATCACTGATGATACATATAAATAAATGTGATATATCATACATCAATGAATGTTATT	972		

RESULT. T. A

```

RESUL 4
US-10-205-428-1006/c
; Sequence 1006, Application US/10205428
; Publication No. US20030108907A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAl17C1

```



```
; CURRENT APPLICATION NUMBER: US/10/205,428
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1019
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1006
; LENGTH: 7374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-428-1006

Query Match          7.5%; Score 42.4; DB 15; Length 7374;
Best Local Similarity 57.6%; Pred. No. 1.1; Indels 0; Gaps 0;
Matches 76; Conservative 0; Mismatches 56;

QY 387 TCCTTGAGGGTGAGTGTAGGGCTGAAGACAGAGAGACAGAGACAAATTCAAAT 446
Db 1091 TACATGAATGTTTGTTCAGCATTTATACATAATGACCAAAAAAATGAATATTCAT 1032

QY 447 GTCTTACACAGACATGATTATATAGATAATGGTACATTCATATAACATGATATTT 506
Db 1031 GTCCATCAACTGTAGTATAGATAATAAATGTCATATATCCATCAATGAATGTTATT 972

QY 507 TACTAATTAATAAA 518
Db 971 CAGCAATAAAAA 960

RESULT 5
US-10-311-455-2130/c
; Sequence 2130, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENERBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2130
; LENGTH: 16766
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2130

Query Match          7.5%; Score 42; DB 15; Length 16766;
Best Local Similarity 60.5%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 45;

QY 422 ACAGGAGACAGAGACAAATTCAAATGTCCTTACACAGAGAGACTGATTATATAGATAATGG 481
Db 503 AAAAAAACTAAATAATTCAAATATCTATATAATAAATACTAATTATATAATAATAA 444

QY 482 TACATTCATATAAACATGATATATTTACTAATTAATAAATACCAACACACACAA 535
Db 443 TATATCTATAAAATAAACTATATCCAACTATTTAAAAAACAATAAAAAA 390

RESULT 6
US-10-292-798-1369/c
; Sequence 1369, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-08-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1369
; LENGTH: 744802
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(744802)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(246)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25640)..(25677)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27078)..(27094)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (141192)..(141769)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159571)..(159606)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (174525)..(174575)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234891)..(235013)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (235514)..(235560)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (279677)..(279729)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (408660)..(409123)
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: (409204)..(409669)
FEATURE:
NAME/KEY: CDS
LOCATION: (428381)..(428396)
FEATURE:
NAME/KEY: CDS
LOCATION: (472204)..(472330)
FEATURE:
NAME/KEY: CDS
LOCATION: (714252)..(714355)
FEATURE:
NAME/KEY: CDS
LOCATION: (714447)..(714529)
FEATURE:
NAME/KEY: CDS
LOCATION: (739794)..(739891)
FEATURE:
NAME/KEY: CDS
LOCATION: (744484)..(744602)
FEATURE:
NAME/KEY: modified_base
LOCATION: (51812)..(51911)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (57122)..(57221)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (79368)..(79467)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (293951)..(294050)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (310089)..(310188)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (332935)..(332935)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (332992)..(332992)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (362002)..(362101)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (639781)..(639880)
OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-1369
Query Match 7.5%; Score 42; DB 16; Length 744802;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
Qy 426 GAGCAGAGACAAATTCCTTACACAGAACTGATATATAGATATAGTGTACA 485
Db 609737 GAAAGGAGCAATTCAGATGTCGTCATAGGGACTCTTTGGATAAATAATGCTGCT 609678
Qy 486 TTCATATAACATGATATATTTACTAATTA 515
Db 609677 TGCATCTATTCATGAGATATTTAGATTTA 609648
US-10-027-632-120056/c
Sequence 120056, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 120056
LENGTH: 915
TYPE: DNA
ORGANISM: Human
US-10-027-632-120056
Query Match 7.4%; Score 41.8; DB 13; Length 915;
Best Local Similarity 63.4%; Pred. No. 0.49;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
Qy 418 AAGAACAGGACAGACAGCAATTCAAATGTCCTTACACAGAACTGATATATAGATA 477
Db 577 AATACCCCAACACAGTAAACAATCCAAAGACCCATTCACAGGTGAAGAATAAATAATT 518
Qy 478 ATGGTACATTCATATAAATCATGATATATTTACTAATTAATAA 518
Db 517 AGCTATATTCATACATAGATATCTATCTCACAATTAAGA 477
RESULT 8
US-10-027-632-120056/c
Sequence 120056, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 120056
```

; LENGTH: 915
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-120056

Query Match 7.4%; Score 41.8; DB 16; Length 915;
Best Local Similarity 63.4%; Pred. No. 0.49; 37; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 418 AAGAACAGGAGACAGACAGCAATTCCTTACACAGAGACTGATTATATAGATA 477
Db 577 AATAACCCCAACAGTAAACATCCAAAGACCCATTACAGGTGAAGATAATAAATT 518
QY 478 ATGTCATTCATATACATGATATATTTACTAATAATAA 518
Db 517 ATGCTATATTCATACATAGATAACTATCTCACAATAAAGA 477

RESULT 9

US-10-085-117-334/c
; Sequence 334, Application US/10085117
; Publication No. US20030232334A1

; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 154817
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: variation
; LOCATION: (1)---(154817)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-334

Query Match 7.4%; Score 41.8; DB 16; Length 154817;
Best Local Similarity 63.4%; Pred. No. 8.6;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 418 AAGAACAGGAGACAGACAGCAATTCCTTACACAGAGACTGATTATATAGATA 477
Db 88945 AACAGCCAAAAGTGGAAACACTCAAATGCTCCATCACTGATGAATCGATAAATT 88886
QY 478 ATGTCATTCATATACATGATATATTTACTAATAATAA 518
Db 88885 GTGGTATATCTATATATGGAATATATTTAGCCATTAATA 88845

RESULT 10

US-09-962-436-281
; Sequence 281, Application US/09962436
; Patent No. US20020081301A1

; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 281
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-281

Query Match 7.3%; Score 41.2; DB 9; Length 167343;
Best Local Similarity 67.4%; Pred. No. 13;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 409 ATAGGGCTGAAGAACAGGAGACAGAGCAATTCAAATGTCTTACACAGAGACTGATT 468
Db 159637 ATTGTTCAAAATACAAAAGATTGGAAGAAGGCAATATCTTCAGTAGAAGACTGATG 159696
QY 469 ATATAGATAATGGTACATTCATATAA 494
Db 159697 AAATACATTGTGCTACATACATACAA 159722

RESULT 11

US-09-964-824A-273
; Sequence 273, Application US/09964824A
; Patent No. US20020102531A1

; GENERAL INFORMATION:
; APPLICANT: Horrigan, Stephen
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signat
; FILE REFERENCE: 689290-73
; CURRENT APPLICATION NUMBER: US/09/964,824A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/60/236,033
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,032
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,028
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 583
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 273
; LENGTH: 167343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-824A-273

Query Match 7.3%; Score 41.2; DB 9; Length 167343;
Best Local Similarity 67.4%; Pred. No. 13;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 409 ATAGGGCTGAAGAACAGGAGACAGAGCAATTCAAATGTCTTACACAGAGACTGATT 468
Db 159637 ATTGTTCAAAATACAAAAGATTGGAAGAAGGCAATATCTTCAGTAGAAGACTGATG 159696
QY 469 ATATAGATAATGGTACATTCATATAA 494
Db 159697 AAATACATTGTGCTACATACATACAA 159722

RESULT 12

US-10-322-281-287
; Sequence 287, Application US/10322281
; Publication No. US20040126762A1

; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287
; LENGTH: 97415
; TYPE: DNA

```
; ORGANISM: Mus. musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(97415)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-287

Query Match          7.3%; Score 41; DB 17; Length 97415;
Best Local Similarity 61.9%; Pred. No. 11;
Matches 65; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 416 TGAGACAGACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATTATATAGA 475
Db 48437 TTAATGCCAGAGCTGGAACACCCAGAAATTCCTCTCAACAGAAAGTGGATACAAAA 48496

Qy 476 TAATGGTACATTCATATAAACAATGATATATTTACTTAATTAACA 520
Db 48497 ATAATGGTACATTTACACAATGGAGTACTCTTGTATTAACA 48541

RESULT 13
US-10-085-783A-16076/c
; Sequence 16076, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16076
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-16076

Query Match          7.1%; Score 40.2; DB 13; Length 414;
Best Local Similarity 59.0%; Pred. No. 0.88;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 409 ATAGGGCTGAAGAACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATT 468
Db 239 ATTATGCTTAACAGCCCCCAACTGGAATTAATTCAAATGTCCTTACATTAAGATGGAT 180

Qy 469 ATATAGATAATGGTACATTCATATAAACAATGATATATTTACTTAATTAACAATACCA 525
Db 179 GAATAAATTCCTAGTATATTCATACAAATGCAAGAGCAATGAAATGAATCCTCTGCTA 123

RESULT 14
US-10-242-535A-16076/c
; Sequence 16076, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
```

```
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16076
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-16076

Query Match          7.1%; Score 40.2; DB 16; Length 414;
Best Local Similarity 59.0%; Pred. No. 0.88;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 409 ATAGGGCTGAAGAACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATT 468
Db 239 ATTATGCTTAACAGCCCCCAACTGGAATTAATTCAAATGTCCTTACATTAAGATGGAT 180

Qy 469 ATATAGATAATGGTACATTCATATAAACAATGATATATTTACTTAATTAACAATACCA 525
Db 179 GAATAAATTCCTAGTATATTCATACAAATGCAAGAGCAATGAAATGAATCCTCTGCTA 123

RESULT 15
US-10-240-425-622
; Sequence 622, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Bolland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 622
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AI425019
US-10-240-425-622

Query Match          7.1%; Score 40.2; DB 13; Length 509;
Best Local Similarity 59.0%; Pred. No. 0.98;
Matches 69; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 409 ATAGGGCTGAAGAACAGACAGACAGACAAATTCAAATGTCCTTACACAGAACTGATT 468
Db 263 ATTATGCTTAACAGCCCCCAACTGGAATTAATTCAAATGTCCTTACATTAAGATGGAT 322

Qy 469 ATATAGATAATGGTACATTCATATAAACAATGATATATTTACTTAATTAACAATACCA 525
Db 323 GAATAAATTCCTAGTATATTCATACAAATGCAAGAGCAATGAAATGAATCCTCTGCTA 379

Search completed: August 15, 2004, 01:59:54
Job time : 322.797 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: August 14, 2004, 11:00:56 ; Search time 54.0891 Seconds
(without alignments)
5776.350 Million cell updates/sec

Title: US-10-082-830-99

Perfect score: 563

Sequence: 1 ggtacatctgctgtgat.....gccgctagctcagctag 563

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Databases :

Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	53.2	9.4	7218	1	US-08-232-463-14
C 2	39	6.9	81001	4	US-09-750-580-1
C 3	37.4	6.6	408	4	US-09-621-976-8793
C 4	35.8	6.4	5535	4	US-10-204-708-18
C 5	35.6	6.3	112132	4	US-09-741-150-3
C 6	35.6	6.3	112132	4	US-10-160-187-3
C 7	35.4	6.3	3083	2	US-08-480-994-36
C 8	35.4	6.3	3083	2	US-08-616-844-36
C 9	35.4	6.3	3083	2	US-08-599-654-36
C 10	35.4	6.3	3083	2	US-08-485-573-36
C 11	35.4	6.3	3083	3	US-08-944-868A-36
C 12	35.4	6.3	3083	3	US-08-944-423A-36
C 13	35.4	6.3	3083	3	US-08-925-743-36
C 14	35.4	6.3	3083	3	US-08-944-496-36
C 15	35.4	6.3	3083	3	US-08-925-767-36
C 16	35.4	6.3	3084	3	US-08-826-246-11
C 17	35.4	6.3	3084	3	US-08-944-495-11
C 18	35.4	6.3	3084	3	US-09-126-640-6
C 19	35.4	6.3	3084	3	US-08-925-588-11
C 20	35.4	6.3	3084	4	US-09-288-292A-6
C 21	35.4	6.3	3084	4	US-09-372-044-11
C 22	35.4	6.3	3084	4	US-08-825-486-11
C 23	35.4	6.3	6113	4	US-10-204-708-13
C 24	35.4	6.3	29604	3	US-08-781-891-207
C 25	35.4	6.3	29604	4	US-09-618-166-207
C 26	34.8	6.2	6113	4	US-10-204-708-14
C 27	34.6	6.1	11015	4	US-10-204-708-55

C 28	34.2	6.1	832	4	US-09-621-976-2813	Sequence 2813, App
C 29	34.2	6.1	1313	4	US-09-149-476-112	Sequence 112, App
C 30	34.2	6.1	319608	4	US-09-539-333D-1	Sequence 1, Appli
C 31	34.2	6.1	319608	4	US-09-679-409-1	Sequence 1, Appli
C 32	34	6.0	3480	4	US-09-331-359-1	Sequence 39, Appli
C 33	34	6.0	19513	4	US-10-204-708-39	Sequence 3, Appli
C 34	34	6.0	83450	4	US-09-811-469-3	Sequence 290, App
C 35	33.8	6.0	3591	4	US-09-107-532A-290	Sequence 19, Appli
C 36	33.8	6.0	35100	1	US-08-306-891B-19	Sequence 17, Appli
C 37	33.8	6.0	35100	5	PCT-US93-06251-19	Sequence 66, Appli
C 38	33.8	6.0	176373	3	US-09-128-155-17	Sequence 66, Appli
C 39	33.6	6.0	4220	1	US-08-832-883-66	Sequence 352, App
C 40	33.6	6.0	4220	2	US-08-832-877-66	Sequence 12, Appli
C 41	33.6	6.0	9956	4	US-08-956-171E-352	Sequence 38, Appli
C 42	33.6	6.0	18073	3	US-09-078-294-12	Sequence 11, Appli
C 43	33.6	6.0	128779	4	US-09-497-855A-38	Sequence 1, Appli
C 44	33.6	6.0	392000	4	US-10-027-993-11	Sequence 1, Appli
C 45	33.4	5.9	4084	2	US-08-568-459A-1	

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/POCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpc-Fls
US-08-232-463-14

Query Match

9.4%; Score 53.2; DB 1; Length 7218;


```
Db      2362 TTAAATC 2356
Query Match      6.3%; Score 35.6; DB 4; Length 112132;
Best Local Similarity 54.6%; Pred. No. 6.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

RESULT 5
US-09-741-150-3/c
; Sequence 3, Application US/09741150
; Patent No. 6436689
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000968
; CURRENT APPLICATION NUMBER: US/09/741.150
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 112132
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112132)
; OTHER INFORMATION: n = A,T,C or G
US-09-741-150-3

Query Match      6.3%; Score 35.6; DB 4; Length 112132;
Best Local Similarity 54.6%; Pred. No. 6.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY      402 GAGAGTGATAGGGCTGAAGAACAGGACAGACAGCAATTCAAATGCTCTTACACAGAAG 461
Db      24899 GACAGCTTTATTTTAAATAGCCAAAACCTGAAACCAACCAAAATGCTATCAACAGATG 24840

QY      462 ACTGATTATATAGATGATGTTACATTCATATATACATGATATATTTACTAATTAACAAT 521
Db      24839 AATTGATAAACAATGATGGTATATCCATACACAGATACACTCCAGCAATAAAAAGA 24780

QY      522 ACCAACACAC 531
Db      24779 AACAAATTAC 24770

RESULT 6
US-08-480-994-36
; Sequence 36, Application US/08480994
; Patent No. 5834248
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,994
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-033
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-480-994-36

Db      2362 TTAAATC 2356
Query Match      6.3%; Score 35.6; DB 4; Length 112132;
Best Local Similarity 54.6%; Pred. No. 6.2;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY      402 GAGAGTGATAGGGCTGAAGAACAGGACAGACAGCAATTCAAATGCTCTTACACAGAAG 461
Db      24899 GACAGCTTTATTTTAAATAGCCAAAACCTGAAACCAACCAAAATGCTATCAACAGATG 24840

QY      462 ACTGATTATATAGATGATGTTACATTCATATATACATGATATATTTACTAATTAACAAT 521
Db      24839 AATTGATAAACAATGATGGTATATCCATACACAGATACACTCCAGCAATAAAAAGA 24780

QY      522 ACCAACACAC 531
Db      24779 AACAAATTAC 24770

RESULT 6
US-10-160-187-3/c
; Sequence 3, Application US/10160187
; Patent No. 6620607
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al.
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000968DIV
; CURRENT APPLICATION NUMBER: US/10/160,187
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: 60/252,410
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: 09/741,150
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 112132
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112132)
; OTHER INFORMATION: n = A,T,C or G
US-10-160-187-3
```


; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-599-654-36
Query Match 6.3%; Score 35.4; DB 2; Length 3083;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 418 AAGACAGGACAGACAGACAAATTCAAATGTCCTTACACAGAGAGACTGATTATATAGATA 477
Db 821 AACAGCCAAAGAGAGAGAGCAACCCAAATGTCCATTAGCTGATAATGGAATGAAT 880
QY 478 ATGGTACATTCATATAACATGATATATTTACTAATTAATAAAACATACCAACACACACAACA 537
Db 861 ATGGTACGTCGAGAGATGGAATATCATTACCCATGAAAAGAACGAAGTCCAGCACCA 940
QY 538 CCTCGAGCGGCTA 550
Db 941 AACGTGCTACAA 953
RESULT 11
US-08-944-868A-36
; Sequence 36, Application US/08944868A
; Patent No. 6018025
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,868A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,654
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:

; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-599-654-36
Query Match 6.3%; Score 35.4; DB 2; Length 3083;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 418 AAGACAGGACAGACAGACAAATTCAAATGTCCTTACACAGAGAGACTGATTATATAGATA 477
Db 821 AACAGCCAAAGAGAGAGAGCAACCCAAATGTCCATTAGCTGATAATGGAATGAAT 880
QY 478 ATGGTACATTCATATAACATGATATATTTACTAATTAATAAAACATACCAACACACACAACA 537
Db 861 ATGGTACGTCGAGAGATGGAATATCATTACCCATGAAAAGAACGAAGTCCAGCACCA 940
QY 538 CCTCGAGCGGCTA 550
Db 941 AACGTGCTACAA 953
RESULT 10
US-08-485-573-36
; Sequence 36, Application US/08485573
; Patent No. 5968770
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,573
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:

```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; US-08-944-868A-36

Query Match
Best Local Similarity 6.3%; Score 35.4; DB 3; Length 3083;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACACATTCAAATGCTTACACAGAGACTGATTATATAGATA 477
Db 821 AACACCCAAAGAGAGAGACACCCAAATGTCATTAGCTGATAAATGATAAATGAAT 880

QY 478 ATGGTACATTTCATATAACATGATATATTTACTAATTAATTAACATACCAACACACACA 537
Db 881 ATGGTACGTCGAGAGATGGAATATCATCCCATGAAAGACGAGTCCAGCACCA 940

QY 538 CCTCAGCGCGCTA 550
Db 941 AAACGTGCTACAA 953

RESULT 12
US-08-944-423A-36
; Sequence 36, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
```

```
;
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
; US-08-944-423A-36

Query Match
Best Local Similarity 6.3%; Score 35.4; DB 3; Length 3083;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACACATTCAAATGCTTACACAGAGACTGATTATATAGATA 477
Db 821 AACACCCAAAGAGAGAGACACCCAAATGTCATTAGCTGATAAATGATAAATGAAT 880

QY 478 ATGGTACATTTCATATAACATGATATATTTACTAATTAATTAACATACCAACACACA 537
Db 881 ATGGTACGTCGAGAGATGGAATATCATCCCATGAAAGACGAGTCCAGCACCA 940

QY 538 CCTCAGCGCGCTA 550
Db 941 AAACGTGCTACAA 953

RESULT 13
US-08-925-743-36
; Sequence 36, Application US/08925743
; Patent No. 6054558
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,573
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
```

```

; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-925-743-36

Query Match 6.3%; Score 35.4; DB 3; Length 3083;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACAAATTCAAATGTCTTACACAGAGACTGATTATATAGATA 477
DB 821 AACGCCAAAGAGAGGAGCAACCAATGTCCATTAGCTGATAAATGGATAAATGAAT 880
QY 478 ATGGTACATTCATATAACATGATATATTTACTAATTAACATACCAACACACACACA 537
DB 881 ATGGTACGTCCGAGATGGAATATCATTCACCCATGAAAAAGAAACGAAGTCCAGCACCA 940
QY 538 CCTCGAGCGGCTA 550
DB 941 AAACGTGCTACAA 953

RESULT 14
US-08-944-496-36
; Sequence 36, Application US/08944496
; Patent No. 6124433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,496
; FILING DATE: 06-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA

; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-944-496-36

Query Match 6.3%; Score 35.4; DB 3; Length 3083;
Best Local Similarity 54.1%; Pred. No. 1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 418 AAGACAGGAGACAGACAAATTCAAATGTCTTACACAGAGACTGATTATATAGATA 477
DB 821 AACGCCAAAGAGAGGAGCAACCAATGTCCATTAGCTGATAAATGGATAAATGAAT 880
QY 478 ATGGTACATTCATATAACATGATATATTTACTAATTAACATACCAACACACACACA 537
DB 881 ATGGTACGTCCGAGATGGAATATCATTCACCCATGAAAAAGAAACGAAGTCCAGCACCA 940
QY 538 CCTCGAGCGGCTA 550
DB 941 AAACGTGCTACAA 953

RESULT 15
US-08-925-767-36
; Sequence 36, Application US/08925767
; Patent No. 6225084
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,767
; FILING DATE: 09-SEPT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
```

```
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2911
US-08-925-767-36

Query Match      6.3%; Score 35.4; DB 3; Length 3083;
Best Local Similarity 54.1%; Pred.No.1.8;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 418 AAGACAGGAGACAGACAAATTCAAATGTCCTTACACAGAGACTGATTATAGATA 477
Db 821 AACAGCCAAAGAGAGGAGCAACCCAAATGTCATTAGCTGATAAATGGATAAATGAAT 880
Qy 478 ATGGTACATTTCATATAAACATGATATATTTACTAATTAACATACCAACACACACAACA 537
Db 881 ATGGTACGTCGGAAGATGGAATATCATTCACCCATGAAAAGAACCGAAGTCCAGCACCA 940
Qy 538 CCTCGAGCCGCTA 550
Db 941 AAACGTGCTACAA 953
```

Search completed: August 14, 2004, 20:48:11
Job time : 56.0891 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:12:36 ; Search time 16 seconds
(without alignments)

288.575 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFTLRQSFLLSQGVAMH.....LPGSSDSRASQASARITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	144	58.1	100	2 A46010	X-linked retinopat
2	103	41.5	613	4 C40201	artifact-warning s
3	83.5	33.7	673	4 F40201	artifact-warning s
4	83	33.5	627	4 A40201	artifact-warning s
5	79.5	32.1	407	2 T02670	probable thromboxa
6	79	31.9	79	2 A56194	thromboxane A-2 re
7	74.5	30.0	597	4 B40201	artifact-warning s
8	67	27.0	1156	2 A47397	aducan homolog -
9	62	25.0	499	2 S65657	alpha-1C-adrenergi
10	61	24.6	522	2 T08711	gamma-adaptin homo
11	61	24.6	726	1 UYPV51	noncapsid protein
12	59	23.8	440	2 JC5520	serotonin receptor
13	59	23.8	841	1 I78895	serine/threonine-s
14	58	23.4	244	2 A42704	methyltransferase
15	58	23.4	244	2 H97486	biotin synthetase p
16	58	23.4	407	2 I52703	42k membrane glyco
17	57.5	23.2	313	2 T06147	hypothetical prote
18	57.5	23.2	1127	2 T32404	hypothetical prote
19	57	23.0	811	2 S76690	hypothetical prote
20	56.5	22.8	997	2 A87330	TonB-dependent rec
21	56	22.6	412	2 S76239	hypothetical prote
22	56	22.6	840	2 A12153	hypothetical prote
23	56	22.6	957	2 H69141	hypothetical prote
24	55	22.2	46	2 I54375	gene NF2 protein -
25	55	22.2	1386	2 T00257	hypothetical prote
26	55	22.2	2326	2 B47447	calcium channel pr
27	55	22.2	2774	2 A43359	microtubule-associ
28	54.5	22.0	530	2 JC7979	cellulobiohydrolase
29	54.5	22.0	547	2 T36550	hypothetical prote

30 54 21.8 1459 2 T30196 kinesin motor prot
31 53.5 21.6 116 2 E72786 hypothetical prote
32 53.5 21.6 167 2 S56544 fibrial protein f
33 53.5 21.6 167 2 F92288 fibrial morpholog
34 53.5 21.6 167 2 A86130 fibrial morpholog
35 53.5 21.6 205 2 T51052 Arp2/3 complex sub
36 53.5 21.6 301 4 B40201 artifact-warning s
37 53 21.4 39 2 I54374 gene NF2 protein -
38 53 21.4 274 2 D97653 hypothetical prote
39 53 21.4 274 2 AB2877 conserved hypothet
40 53 21.4 436 2 E71493 probable RNA polym
41 53 21.4 653 2 A46362 amyloid precursor-
42 53 21.4 779 2 H71301 probable membrane-
43 52.5 21.2 53 2 A42442 integrin beta-1 ch
44 52.5 21.2 254 2 F69547 nucleotide-binding
45 52.5 21.2 348 1 MNHHPB outer membrane pro

ALIGNMENTS

RESULT 1

A46010
X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C;Accession: A46010
R;Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.
Genomics 15, 467-471, 1993
A;Title: Identification and partial characterization of a candidate gene for X-linked r
A;Reference number: A46010; MUID:93224131; PMID:8468040
A;Accession: A46010
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-100 <WON>
A;Cross-references: GB:S58722; NID:g299470; PIDN:AAB26149.1; PID:g299471
A;Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBI:129340)

Query Match 58.1%; Score 144; DB 2; Length 100;

Best Local Similarity 62.5%; Pred. No. 2.2e-11;

Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFTLRQSFLLSQGVAMHDLGSLHPPLPGSSDSRASQASARITGV 48

Db 2 FFFFTETSCVAEAGVQWCDLGLSKSPPPGSSDSPASASRVAGITGM 49

RESULT 2

C40201
artifact-warning sequence (translated ALU class C) - human

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C;Accession: C40201

R;Claverie, J.M.

A;Reference number: A40201

A;Accession: C40201

A;Molecule type: DNA

A;Residues: 1-613 <CLA>

R;Claverie, J.M.

Genomics 12, 838-841, 1992

A;Title: Identifying coding exons by similarity search: Alu-derived and other potential

A;Reference number: A40200; MUID:92241891; PMID:1572661

A;Contents: annotation

C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'X'.

C;Comment: Any significant similarity of a predicted protein sequence to a portion of t

Query Match

Best Local Similarity 61.3%; Score 103; DB 4; Length 613;

Matches 19; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQGVAVHDLGSLHPLPG 31
 ||||| : : : : : ||||| : : ||
 DB 311 FFFFEESHVTOAGVQWRDLGSLQAPPG 341

RESULT 3

F40201
 artifact-warning sequence (translated ALU class F) - human

C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 C:Accession: F40201

R:Claverie, J.M.
 personal communication, 1992

A:Reference number: A40201

A:Accession: F40201

A:Molecule type: DNA

A:Residues: 1-673 <CLA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 33.7%; Score 83.5; DB 4; Length 673;
 Best Local Similarity 43.6%; Pred. No. 0.01; Mismatches 21; Indels 7; Gaps 1;
 Matches 24; Conservative 3;

QY 1 FFFFLRQSFLLSQGVAVHDLGSLHPLPGSSDRASASQSRITGV 48

DB 344 FFFFCFVFETGPCSDTGAQVQWVLAHCSNLGLGSSDSPASVSRVAGITGM 398

RESULT 4

A40201
 artifact-warning sequence (translated ALU class A) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
 C:Accession: A40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: A40201

A:Molecule type: DNA

A:Residues: 1-627 <CLA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 33.5%; Score 83; DB 4; Length 627;
 Best Local Similarity 51.6%; Pred. No. 0.011;
 Matches 16; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQGVAVHDLGSLHPLPG 31

DB 321 FLFFPSETSRVAKAGVQWRDLGSLQAPPG 351

RESULT 5

T02670

probable thromboxane A2 receptor isoform beta - human

C:Species: Homo sapiens (man)

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 15-Sep-2000
 C:Accession: T02670

R:Lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; G
 J.; Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; At
 P.; Quan, G.; Krommiller, B.; Arellano, A.; Montgomerly, M.; Ow, D.; Nolan, M.
 submitted to the EMBL Data Library, June 1998

A:Authors: Trong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.

A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine
 A:Reference number: Z14696

A:Accession: T02670

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-407 <LAM>

A:Cross-references: EMBL:AC005175; NID:G3253116; PID:G3253117

C:Superfamily: prostaglandin E receptor EPI

Query Match 32.1%; Score 79.5; DB 2; Length 407;

Best Local Similarity 49.0%; Pred. No. 0.019; 6; Mismatches 13; Gaps 3;
 Matches 24; Conservative 6;

QY 7 ROSFTLSQGVAVHDL---GSL---HPPLFGSSDSRASASQSRITGV 48

DB 328 RESLTL-----WPSLEYSGTISAHCNLRPLPGSSDSRASASRAAGITGV 370

RESULT 6

A56194

thromboxane A-2 receptor, endothelial - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 07-May-1999

C:Accession: A56194

R:Raychowdhury, M.K.; Yukawa, M.; Collins, L.J.; McGrail, S.H.; Kent, K.C.; Ware, J.A.

J. Biol. Chem. 270, 7011, 1995

A:Reference number: A56194; MUID:95204505; PMID:7896853

A:Accession: A56194

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-73 <RAY>

A:Cross-references: GB:U11271

C:Superfamily: prostaglandin E receptor EPI

Query Match 31.9%; Score 79; DB 2; Length 79;

Best Local Similarity 85.0%; Pred. No. 0.0038;
 Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 29 LPGSSDSRASASQSRITGV 48

DB 23 LPGSSDSRASASRAAGITGV 42

RESULT 7

E40201

artifact-warning sequence (translated ALU class E) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000

C:Accession: E40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: E40201

A:Molecule type: DNA

A:Residues: 1-597 <CLA>

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 30.0%; Score 74.5; DB 4; Length 597;

Best Local Similarity 71.4%; Pred. No. 0.13; 2; Indels 1; Gaps 1;
 Matches 15; Conservative 3; Mismatches 1;

QY 3 FFFLRQSTLL-SQAGVAHDL 22
 |||||:| :|||:|
 Db 501 FFFLRSPALVAQAGVRWHL 521

RESULT 8

A47397
 aducin homolog - fruit fly (Drosophila melanogaster)
 N/Alternate names: D-add; gene hu-li tai shao protein; membrane-cytoskeleton-associated
 C/Species: Drosophila melanogaster
 C/Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C/Accession: A47397; A46392
 R/Ding, D.; Parkhurst, S.M.; Lipshitz, H.D.
 Proc Natl Acad Sci U S A 90, 2512-2516, 1993
 A/Title: Different genetic requirements for anterior RNA localization revealed by the dh
 A/Reference number: A47397; MUID:93211992; PMID:7681599
 A/Accession: A47397
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-1156 <DIN>
 A/Cross-references: GB:L14330; GB:L07617; NID:9290210; PIDN:BAE59182.1; PID:G290211
 A/Experimental source: embryo
 A/Note: sequence extracted from NCBI backbone (NCBIN:128276, NCBI:P:128277)
 R/Yue, L.; Spradling, A.C.
 Genes Dev. 6, 2443-2454, 1992
 A/Title: hu-li tai shao, a gene required for ring canal formation during Drosophila ooge
 A/Reference number: A46392; MUID:94040709; PMID:1340461
 A/Accession: A46392
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-350, 'V', 352-679, 'G', 681, 'L', 683-724, 'L', 726, 'H', 728-745, 'F', 747-911, 'D', 91
 A/Cross-references: GB:L05016; NID:G157746; PID:G157747
 A/Experimental source: egg
 C/Comment: This protein is required for assembling actin at ring canals in developing eg
 C/Genetics:
 A/Gene: FlyBase:hts; hu-li tai shao
 A/Cross-references: FlyBase:FBgn0004873
 C/Keywords: egg; embryo

Query Match 27.0%; Score 67; DB 2; Length 1156;

Best Local Similarity 39.5%; Pred. No. 2.4; Mismatches 18; Indels 0; Gaps 0;

Matches 15; Conservative 5; Mismatches 18; Indels 0; Gaps 0;

QY 6 LRQSTLLSQAGVAHDLGSLHPPPGSSDSRASASQSA 43

Db 1039 LHEQQTGRAPLTPSFTHPPAPASSSMVHRNSA 1076

RESULT 9

S65657
 alpha-1C-adrenergic receptor splice form 2 - human
 N/Alternate names: alpha-1C-adrenoceptor isoform 2
 C/Species: Homo sapiens (man)
 C/Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Jun-2000
 C/Accession: S65657; S65655
 R/Tanaka, T.,
 submitted to the EMBL Data Library, July 1994
 A/Reference number: S65656
 A/Accession: S65657
 A/Molecule type: mRNA
 A/Residues: 1-499 <TAN>
 A/Cross-references: EMBL:D32202; NID:9272208; PIDN:BA06901.1; PID:G927209
 R/Hirasawa, A.; Shibata, K.; Horie, K.; Takei, Y.; Obika, K.; Tanaka, T.; Muramoto, N.;
 FEBS Lett. 363, 256-260, 1995
 A/Title: Cloning, functional expression and tissue distribution of human alpha(1C)-adren
 A/Reference number: S65654; MUID:95255557; PMID:7737411
 A/Accession: S65655
 A/Molecule type: mRNA
 A/Residues: 424-499 <HIR>
 A/Cross-references: EMBL:D32202
 C/Genetics:
 A/Gene: GDB:ADRA1C; ADRA11

A/Cross-references: GDB:128088; OMIM:104221

A/Map position: 8p21-ep11.2

C/Superfamily: vertebrate rhodopsin

C/Keywords: alternative splicing; neurotransmitter receptor

Query Match 25.0%; Score 62; DB 2; Length 499;

Best Local Similarity 73.7%; Pred. No. 4.3;

Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 29 LQSSDSRASASQSARITG 47

Db 446 LQGRDSPASASQAGTIG 464

RESULT 10

T08711
 gamma-adaptin homolog DKFp564D066.1 - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
 C/Accession: T08711
 R/Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, March 1999
 A/Reference number: Z16471
 A/Accession: T08711
 A/Molecule type: mRNA
 A/Residues: 1-522 <WAM>
 A/Cross-references: EMBL:AL050025
 A/Experimental source: fetal brain; clone DKFp564D066
 C/Genetics:
 A/Note: DKFp564D066.1
 C/Keywords: membrane trafficking

Query Match 24.6%; Score 61; DB 2; Length 522;

Best Local Similarity 56.0%; Pred. No. 6.1;

Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 6 LRQSTLLSQAGVAHDLGSLHPPPLP 30

Db 495 LTGSHSVSQAGVQWDYLGSLQLPFP 519

RESULT 11

UYPVSL

noncapsid protein NS1 - bovine parvovirus

C/Species: bovine parvovirus

C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999

C/Accession: C26104

R/Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.

J. Virol. 60, 1085-1097, 1986

A/Title: Complete nucleotide sequence and genome organization of bovine parvovirus.

A/Reference number: A26104; MUID:87061194; PMID:3783814

A/Accession: C26104

A/Molecule type: DNA

A/Residues: 1-726 <CHE>

A/Cross-references: EMBL:M14363; NID:G333454; PIDN:AAE59845.1; PID:G808803; EMBL:M1972

C/Superfamily: parvovirus noncapsid protein

C/Keywords: noncapsid protein

Query Match 24.6%; Score 61; DB 1; Length 726;

Best Local Similarity 30.0%; Pred. No. 8.7;

Matches 12; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 8 QSTLLSQAGVAHDLGSLHPPPGSSDSRASASQSARITG 47

Db 486 QNWTFSENGVCHCGFIQTFESDTSDSGDPPDPAVAG 525

RESULT 12

JC5520

serotonin receptor 6 - human

N/Alternate names: 5-hydroxytryptamine receptor 6 (5-HT6)

C/Species: Homo sapiens (man)

C/Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 05-Nov-1999

C;Accession: JG5520
R.Kohan, R.; Mccall, M.A.; Khan, N.; Druck, T.; Huebner, K.; Lachowicz, J.E.; Meltzer, J. Neurochem. 66, 47-56, 1996
A;Title: Cloning, characterization, and chromosomal localization of a human 5-HT 6 serotonergic receptor
A;Reference number: JG5520; MUID:96102917; PMID:8542988
A;Accession: JG5520
A;Molecule type: mRNA
A;Residues: 1-440 <KOR>
A;Cross-references: GB:L11147; NID:g1162923; PIDN:AAA92622.1; PID:g1162924
A;Experimental source: brain
C;Comment: This protein shows high affinity for several therapeutically important antidiabetic drugs
C;Genetics:
A;Gene: HTR6
A;Map position: 1p35-36
A;Introns: 238/3: 291/3
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; transmembrane protein
F:28-51/Domain: transmembrane #status predicted <TM1>
F:65-88/Domain: transmembrane #status predicted <TM2>
F:100-123/Domain: transmembrane #status predicted <TM3>
F:143-166/Domain: transmembrane #status predicted <TM4>
F:185-208/Domain: transmembrane #status predicted <TM5>
F:266-289/Domain: transmembrane #status predicted <TM6>
F:298-321/Domain: transmembrane #status predicted <TM7>
F:10/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.8%; Score 59; DB 2; Length 440;
Best Local Similarity 45.5%; Pred. No. 9.1;
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 17 VAWHDLGSLHPLPGSSDSRAS 39
DB 164 LGWELGHARPPVPGQCLLAS 185

RESULT 13
I78885
serine/threonine-specific protein kinase (EC 2.7.1.-) STK2 - human
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 29-Sep-1999
C;Accession: I78885
R;Levedakou, E.N.; He, M.; Baptist, E.W.; Craven, R.J.; Cance, W.G.; Walczak, P.L.; Simmonds, J.; et al. J. Biol. Chem. 273, 1977-1988, 1998
A;Title: Two novel human serine/threonine kinases with homologies to the cell cycle regulatory protein p34^{cdc2}
A;Reference number: I58396; MUID:94268838; PMID:9208544
A;Accession: I78885
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-841 <RES>
A;Cross-references: GB:L20321; NID:g348244; PIDN:AAA36658.1; PID:g348245
C;Genetics:
A;Gene: GDB:STK2
A;Cross-references: GDB:374125
A;Map position: 3p21.1-3p21.1
C;Superfamily: human serine/threonine-specific protein kinase homology
C;Keywords: phosphotransferase
F:4-261/Domain: protein kinase homology <KIN>

Query Match 23.8%; Score 59; DB 1; Length 841;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 29 LPGSSDSRASASQSRITGV 48
DB 478 LLGSSDSPASASRVAGITGV 497

RESULT 14
AH2704
methyltransferase Atu1041 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AH2704

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; et al. Science 294, 2317-2323, 2001
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AH2704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-244 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42054.1; PID:g17739432; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1041
A;Map position: circular chromosome

Query Match 23.4%; Score 58; DB 2; Length 244;
Best Local Similarity 37.1%; Pred. No. 6.5;
Matches 23; Conservative 5; Mismatches 16; Indels 18; Gaps 4;

QY 4 FTLRQSFT-----LSQAGVAWHDLGSLHPLPGSS-----DSRASASQ-SARIT 46
DB 11 FFERYSAMPRSTIEGLRQAG-EWHELRLAMPDLKGRSFLDLGCGFGWHCRVAEQGAARIV 69

QY 47 GV 48
DB 70 GV 71

RESULT 15
H97486
biotin synthesis protein bioC VC1114 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: H97486
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Williams, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; et al. Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H97486
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-244 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86849.1; PID:g15156063; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C 1920
A;Map position: circular chromosome

Query Match 23.4%; Score 58; DB 2; Length 244;
Best Local Similarity 37.1%; Pred. No. 6.5;
Matches 23; Conservative 5; Mismatches 16; Indels 18; Gaps 4;

QY 4 FTLRQSFT-----LSQAGVAWHDLGSLHPLPGSS-----DSRASASQ-SARIT 46
DB 11 FFERYSAMPRSTIEGLRQAG-EWHELRLAMPDLKGRSFLDLGCGFGWHCRVAEQGAARIV 69

QY 47 GV 48
DB 70 GV 71

Search completed: August 10, 2004, 19:15:42
Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004, Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:11:21 ; Search time 13 Seconds
(without alignments)

192.259 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLRQSFILSQGVAVH.....LPQSSDSRASQSRITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	43.5	591	ALU8_HUMAN	P39185 homo sapien
2	103.5	41.7	593	ALU7_HUMAN	P39184 homo sapien
3	97	39.1	603	ALU4_HUMAN	P39151 homo sapien
4	96.5	38.9	593	ALU6_HUMAN	P39183 homo sapien
5	83	33.5	587	ALU2_HUMAN	P39189 homo sapien
6	83	33.5	587	ALU3_HUMAN	P39190 homo sapien
7	77	31.0	585	ALU5_HUMAN	P39182 homo sapien
8	76.5	30.8	591	ALU1_HUMAN	P39188 homo sapien
9	73	29.4	438	SCCE_HUMAN	Q43556 homo sapien
10	67	27.0	1156	HTS_DROME	Q02645 drosophila
11	65.5	26.4	741	RED1_HUMAN	P78563 homo sapien
12	65	26.2	211	BCS4_HUMAN	Q8tdm0 homo sapien
13	63	25.4	629	2195_HUMAN	O14628 homo sapien
14	62	25.0	833	MAK1_HUMAN	Q32918 homo sapien
15	61	24.6	726	VNCS_PAVBO	P07236 bovine parv
16	59	23.8	440	SH6_HUMAN	P50406 homo sapien
17	59	23.8	841	NEK4_HUMAN	P51957 homo sapien
18	58	23.4	407	ADRM_HUMAN	Q16186 homo sapien
19	57	23.0	811	Y104_SYNY3	P43371 synecocyst
20	56	22.6	902	GLR4_HUMAN	P48058 homo sapien
21	55	22.2	2326	CCAB_DISOM	P56698 discopys o
22	55	22.2	2774	NAPA_RAT	P34926 rattus norv
23	54	21.8	2333	PGCA_CANFA	Q28343 canis famil
24	54	21.8	2805	NAPA_HUMAN	P78559 homo sapien
25	53.5	21.6	167	FTMG_ECOLI	P08190 escherichia
26	53.5	21.6	271	S24D_ANOGA	Q17004 anopheles g
27	53.5	21.6	847	NAT3_HUMAN	P43243 homo sapien
28	53	21.4	176	NL12_PARJU	Q04404 parietaria
29	53	21.4	653	APPI_MOUSE	Q03157 mus musculu
30	53	21.4	977	RBMF_HUMAN	Q96t37 homo sapien
31	52.5	21.2	348	OMB1_NEIGO	P18195 neisseria g
32	52.5	21.2	350	MRP_DEIRA	Q9vm9 deinococcus
33	52.5	21.2	402	B4G1_BOVIN	P08037 b beta-1,4-

34 52 21.0 407 1 ADRM_MOUSE
35 52 21.0 407 1 ADRM_RAT
36 52 21.0 476 1 EGR2_HUMAN
37 52 21.0 663 1 GLI3_CHICK
38 52 21.0 710 1 IRL1_MOUSE
39 52 21.0 769 1 TFR1_FELCA
40 52 21.0 1011 1 M3K6_HUMAN
41 52 21.0 1183 1 DRPL_RAT
42 52 21.0 1235 1 CYA4_TRYBB
43 51.5 20.8 382 1 Y894_HAEIN
44 51.5 20.8 487 1 WRK1_ARATH
45 51 20.6 226 1 IBP6_RAT

ALIGNMENTS

RESULT 1
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human)
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
pollution of protein sequence databases with Alu-derived amino
acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
primate genomes with an average spacing of 4 kb. Some of them are
actively transcribed by pol III. Normal transcripts may contain
Alu-derived sequences in 5' or 3' untranslated regions. However,
cDNA libraries also contain partial and/or rearranged cDNAs
ligated with Alu-derived sequence in any orientation. Although Alu
elements (especially situated on the complementary strand) have a
great potential to create additional/alternative exons,

consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

-!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14574; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 195 FRAME-2.

FT DOMAIN 199 294 FRAME-3.

FT DOMAIN 298 393 FRAME-4.

FT DOMAIN 397 492 FRAME-5.

FT DOMAIN 496 591 FRAME-6.

FT SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Query Match 43.5%; Score 108; DB 1; Length 591;

Best Local Similarity 32.2%; Pred. No. 6.1e-06;

Matches 28; Conservative 6; Mismatches 11; Indels 42; Gaps 1;

QY 4 FFLRQSFTLSQAGVAVHDLGSLHPPPLG----- 31

DB 496 FFEESRSVAQGVQWRDLGSLQPPFPGRFSCSLPSSWDYRPPRPPANFCIFSRDG 555

QY 32 -----SSDSRASQSQARITGV 48

DB 556 VSPCPWGSRTPTDLRXSSQSGAGITGV 582

RESULT 2

ALU7_HUMAN STANDARD; PRT; 593 AA.

AC P39194;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alu subfamily SQ sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_taxid=9606;

RN [1]

RP MEDLINE=92241891; PubMed=1572661;

RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RL Genomics 12:838-841(1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE=88333009; PubMed=3138422;

RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

J. Mol. Evol. 27:194-202(1988).

[4]

RN ALU FAMILIES CLASSIFICATION.

RX MEDLINE=91178915; PubMed=1706781;

RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -!- CAUTION: This Alu entry is provided in order to avoid the further pollution of protein sequence databases with Alu-derived amino acid sequences.

CC -!- CAUTION: Alu repetitive sequences are interspersed in human and primate genomes with an average spacing of 4 kb. Some of them are actively transcribed by pol III. Normal transcripts may contain Alu-derived sequences in 5' or 3' untranslated regions. However, cDNA libraries also contain partial and/or rearranged cDNAs ligated with Alu-derived sequence in any orientation. Although Alu elements (especially situated on the complementary strand) have a great potential to create additional/alternative exons, consideration should be given to the possibility that the presence of an Alu in an open reading frame may have resulted from a cloning artifact or may be due to misinterpretation of sequencing data. This point has been overlooked on several occasions, with the consequence of erroneous Alu-derived amino acid sequences being reported.

CC -!- CAUTION: Any significant similarity of a putative protein sequence with an Alu-translated entry must be taken as a warning that a part of Alu repeat may have been artifactually included in the coding nucleotide sequence.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U14573; -; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 97 FRAME-1.

FT DOMAIN 101 196 FRAME-2.

FT DOMAIN 200 295 FRAME-3.

FT DOMAIN 299 395 FRAME-4.

FT DOMAIN 399 494 FRAME-5.

FT DOMAIN 498 593 FRAME-6.

FT SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;

Query Match 41.7%; Score 103.5; DB 1; Length 593;

Best Local Similarity 72.4%; Pred. No. 2.2e-05;

Matches 21; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 4 FFLRQSFTLSQAGVAVHDLGSLHPPPLG 31

DB 299 FFLRFSFALVAQGVQWRDLGSLQPPFPFG 327

RESULT 3

ALU4_HUMAN STANDARD; PRT; 603 AA.

ID ALU4_HUMAN

AC P39191;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Alu subfamily SB2 sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN (2)
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN (3)
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN (4)
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Miosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC coding nucleotide sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14570; -; NOT_ANNOTATED_CDS.
DR
KW Hypothetical protein.
FT DOMAIN 1 98
FT DOMAIN 102 199
FT DOMAIN 202 300
FT DOMAIN 304 401
FT DOMAIN 405 502
FT DOMAIN 506 603
FT DOMAIN 506 603

```

```

SQ SEQUENCE 603 AA; 65272 MW; B8AAD0AD46BEA114 CRC64;
Query Match 39.1%; Score 97; DB 1; Length 603;
Best Local Similarity 31.2%; Pred. No. 0.00015;
Matches 29; Conservative 4; Mismatches 12; Indels 48; Gaps 2;
QY 4 FFLRSFTLS---QAGVAWHDLGSLHPLPGSS----- 33
| | | | | | | | | | | | | | | | | | | |
Db 304 FFLRSFTLSPPDCGLQWRNLGSLQAPLPGGTFPSCLSPSSWDYRRPPRPPANFLYFX 363
| | | | | | | | | | | | | | | | | | | |
QY 34 -----DSRASASQSAITGV 48
| | | | | | | | | | | | | | | | | | | |
Db 364 XRRGFTLLARMVVISXPHDPPASQSAITGV 396
| | | | | | | | | | | | | | | | | | | |
RESULT 4
ALU6 HUMAN STANDARD; PRT; 593 AA.
ID ALU6_HUMAN
AC F39193;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily SP sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN (2)
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN (3)
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN (4)
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Miosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
CC acid sequences
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC coding nucleotide sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14570; -; NOT_ANNOTATED_CDS.
DR
KW Hypothetical protein.
FT DOMAIN 1 98
FT DOMAIN 102 199
FT DOMAIN 202 300
FT DOMAIN 304 401
FT DOMAIN 405 502
FT DOMAIN 506 603
FT DOMAIN 506 603

```



```

RT  "Alu alert.";
RL  Nature 371:752-752(1994).
RN  [2]
RP  MEDLINE=92241891; PubMed=1572661;
RX  Claverie J.-M.;
RA  "Identifying coding exons by similarity search: alu-derived and other
RT  potentially misleading protein sequences.";
RL  Genomics 12:838-841(1992).
RN  [3]
RX  MEDLINE=92241891; PubMed=1572661;
RA  Claverie J.-M.;
RL  "The Alu family developed through successive waves of fixation
RT  closely connected with primate lineage history.";
RL  J. Mol. Evol. 27:194-202(1988).
RN  [4]
RX  MEDLINE=92241891; PubMed=1572661;
RA  Claverie J.-M.;
RL  "Reconstruction and analysis of human Alu genes.";
RX  J. Mol. Evol. 32:105-121(1991).
CC  -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC  THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC  CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC  FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC  REPEATS.
CC  -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC  CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC  -!- CAUTION: This Alu entry is provided in order to avoid the further
CC  pollution of protein sequence databases with Alu-derived amino
CC  acid sequences.
CC  -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC  primate genomes with an average spacing of 4 kb. Some of them are
CC  actively transcribed by pol III. Normal transcripts may contain
CC  Alu-derived sequences in 5' or 3' untranslated regions. However,
CC  cDNA libraries also contain partial and/or rearranged cDNAs
CC  ligated with Alu-derived sequence in any orientation. Although Alu
CC  elements (especially situated on the complementary strand) have a
CC  great potential to create additional/alternative exons,
CC  consideration should be given to the possibility that the presence
CC  of an Alu in an open reading frame may have resulted from a
CC  cloning artifact or may be due to misinterpretation of sequencing
CC  data. This point has been overlooked on several occasions, with
CC  the consequence of erroneous Alu-derived amino acid sequences
CC  being reported.
CC  -!- CAUTION: Any significant similarity of a putative protein sequence
CC  with an Alu-translated entry must be taken as a warning that a
CC  part of Alu repeat may have been artifactually included in the
CC  coding nucleotide sequence.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; U14569; -; NOT_ANNOTATED_CDS.
CC  -----
CC  KW  Hypothetical protein.
CC  FT  DOMAIN 1 96 FRAME-1.
CC  FT  DOMAIN 100 194 FRAME-2.
CC  FT  DOMAIN 198 292 FRAME-3.
CC  FT  DOMAIN 296 391 FRAME-4.
CC  FT  DOMAIN 395 489 FRAME-5.
CC  FT  DOMAIN 493 587 FRAME-6.
CC  SQ  SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;
CC  Query Match 33.58; Score 83; DB 1; Length 587;
CC  Best Local Similarity 57.18; Pred. No. 0.0082;
CC  Matches 16; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```

```

QY  4 FFIRQSTLSQAGVAMHDLGSLHPLPG 31
DB  493 FFETESRVAQAGVQWRDLGSLQAPP 520

RESULT 7
ID  ALUS_HUMAN STANDARD; PRT; 585 AA.
AC  P39192;
DT  01-FEB-1995 (Rel. 31, Created);
DT  01-FEB-1995 (Rel. 31, Last sequence update);
DT  10-OCT-2003 (Rel. 42, Last annotation update);
DE  Alu subfamily SC sequence contamination warning entry.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX  NCBII_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=95021758; PubMed=7935834;
RA  Claverie J.-M.;
RT  "Alu alert.";
RL  Nature 371:752-752(1994).
RN  [2]
RP  CONCEPT.
RX  MEDLINE=92241891; PubMed=1572661;
RA  Claverie J.-M.;
RT  "Identifying coding exons by similarity search: alu-derived and other
RL  potentially misleading protein sequences.";
RL  Genomics 12:838-841(1992).
RN  [3]
RP  ALU FAMILIES CLASSIFICATION.
RX  MEDLINE=88333009; PubMed=3138422;
RA  Quentlin Y.;
RT  "The Alu family developed through successive waves of fixation
RL  closely connected with primate lineage history.";
RL  J. Mol. Evol. 27:194-202(1988).
RN  [4]
RP  ALU FAMILIES CLASSIFICATION.
RX  MEDLINE=91178815; PubMed=1706781;
RA  Jurka J.;
RT  "Reconstruction and analysis of human Alu genes.";
RL  J. Mol. Evol. 32:105-121(1991).
CC  -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC  THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC  CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC  FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC  REPEATS.
CC  -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC  CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC  -!- CAUTION: This Alu entry is provided in order to avoid the further
CC  pollution of protein sequence databases with Alu-derived amino
CC  acid sequences.
CC  -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC  primate genomes with an average spacing of 4 kb. Some of them are
CC  actively transcribed by pol III. Normal transcripts may contain
CC  Alu-derived sequences in 5' or 3' untranslated regions. However,
CC  cDNA libraries also contain partial and/or rearranged cDNAs
CC  ligated with Alu-derived sequence in any orientation. Although Alu
CC  elements (especially situated on the complementary strand) have a
CC  great potential to create additional/alternative exons,
CC  consideration should be given to the possibility that the presence
CC  of an Alu in an open reading frame may have resulted from a
CC  cloning artifact or may be due to misinterpretation of sequencing
CC  data. This point has been overlooked on several occasions, with
CC  the consequence of erroneous Alu-derived amino acid sequences
CC  being reported.
CC  -!- CAUTION: Any significant similarity of a putative protein sequence
CC  with an Alu-translated entry must be taken as a warning that a
CC  part of Alu repeat may have been artifactually included in the
CC  coding nucleotide sequence.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration

```



```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14571; -; NOT_ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 95
CC FT DOMAIN 99 193
CC FT DOMAIN 197 291
CC FT DOMAIN 295 389
CC FT DOMAIN 393 487
CC FT DOMAIN 491 585
CC FT SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;
SQ
Query Match 31.0%; Score 77; DB 1; Length 585;
Best Local Similarity 77.8%; Pred. No. 0.046;
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 14 QAGVAVHDGLSLHPLPG 31
DB 403 QAGVQWRDLGSLQPPPG 420
RESULT 8
ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
[2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
[3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
[4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178915; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: This Alu entry is provided in order to avoid the further
CC pollution of protein sequence databases with Alu-derived amino
```

```
CC acid sequences.
CC -!- CAUTION: Alu repetitive sequences are interspersed in human and
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC ligated with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -!- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14567; -; NOT_ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 96
CC FT DOMAIN 100 195
CC FT DOMAIN 193 284
CC FT DOMAIN 298 393
CC FT DOMAIN 397 492
CC FT DOMAIN 496 591
CC FT SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;
SQ
Query Match 30.8%; Score 76.5; DB 1; Length 591;
Best Local Similarity 47.9%; Pred. No. 0.054;
Matches 23; Conservative 4; Mismatches 10; Indels 11; Gaps 2;
QY 2 FFFFLRQSF-TLSQAGVAVHDGLSLHPLPGSSDSRASASQASRITGV 48
DB 351 FVFFVETGFHYVQAQGL-----ELLGSSDPPASASQASAGITGV 388
RESULT 9
SGCE_HUMAN STANDARD; PRT; 438 AA.
ID SGCE_HUMAN
AC O43556;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Epsilon-sarcoglycan precursor (Epsilon-SG).
GN SGCE OR ESG.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Nigro V.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA McNally E.M., Ly C.T., Kunkel L.M.;
RL "Human epsilon-sarcoglycan is highly related to alpha-sarcoglycan
RL (adhalin), the limb girdle muscular dystrophy 2D gene.";
RL FEBS Lett. 422:27-32(1998).
```



```

CC CC -!- SIMILARITY: Contains 2 DBEM (double-stranded RNA-binding) domains.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL; U82120; AAB61686.1; -
DR EMBL; U82121; AAB61687.1; -
DR EMBL; X99227; CAA67611.1; -
DR EMBL; X99383; CAA67762.1; -
DR EMBL; U76420; AAC51240.1; -
DR EMBL; U76421; AAC51241.1; -
DR EMBL; U76422; AAC51242.1; -
DR EMBL; AF001042; AAB58300.1; -
DR EMBL; AL163301; CAB90493.1; -
DR HSSP; Q91836; 1D12.
DR Genew; HGNC:226; ADARB1.
DR GK; P78563; -
DR MIM; 601218; -
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0004000; F:adenosine deaminase activity; TAS.
DR GO; GO:0003725; F:double-stranded RNA binding; TAS.
DR GO; GO:0007417; P:central nervous system development; TAS.
DR GO; GO:0006396; P:RNA processing; TAS.
DR InterPro; IPR002466; A:deamin.
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF02137; A:deamin; 1.
DR Pfam; PF00035; dsrm; 2.
DR SMART; SM00552; ADEAMC; 1.
DR SMART; SM00358; DSRM; 2.
DR PROSITE; PS50141; A:DEAMIN EDITASE; 1.
DR PROSITE; PS50137; DS_RBD; 2.
KW mRNA processing; Hydrolyase; Zinc; RNA-binding; Repeat;
KW Alternative splicing.
FT DOMAIN 76 146
FT DRAIN 230 300
FT METAL 394 394
FT ACT_SITE 396 396
FT METAL 451 451
FT METAL 556 556
FT VARSPPLIC 466 505
FT VARSPPLIC 713 741
FT VARSPPLIC 713 741
FT CONFLICT 30 30
FT CONFLICT 423 423
FT CONFLICT 475 475
SQ SEQUENCE 741 AA; 80763 MW; 02B583414DD59C20 CRC64;
Query Match 26.4%; Score 65.5; DB 1; Length 741;
Best Local Similarity 45.2%; Pred. No. 1.7;
Matches 19; Conservative 3; Mismatches 19; Indels 1; Gaps 1;
Qy 6 LRQFTLLSQAGVAMHDLGLSLPPLFG-SSDSRASASQARIT 46
Db 464 LEGRSVYQAGVQVQCNHSLQRPPLGLSLDPLSTFTFGAGTT 505
RESULT 12
ID BC54 HUMAN STANDARD; PRT; 211 AA.
AC Q8TDM0; Q8IX17; Q8NEZ6; Q8TD19; Q9NX13; Q9Y511;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Breast carcinoma amplified sequence 4.
GN BCAS4.
OS Homo sapiens (Human).

```

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC [1]
CC NCBI_TaxID=9606;
CC RP SEQUENCE FROM N.A. (ISOFORM 1), TISSUE SPECIFICITY, AND CHROMOSOMAL
CC TRANSLOCATION WITH BCAS3.
CC TISSUE=Liver;
CC RX MEDLINE=22265331; PubMed=12378525;
CC RA Baerlund M., Morni O., Weaver J.D., Kauraniemi P., Sauter G.,
CC Ra Heiskanen M., Kallioniemi O.-P., Kallioniemi A.;
CC RT "Cloning of BCAS3 (17q23) and BCAS4 (20q13) genes that undergo
CC RT amplification, overexpression, and fusion in breast cancer.";
CC EL Genes Chromosomes Cancer 35:311-317(2002).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=21638749; PubMed=11780052;
CC RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
CC RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
CC RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,
CC RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
CC RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
CC RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
CC RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
CC RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
CC RA Ellington A.G., Frankland J.A., Fraser A., Griffiths M.N.D., Gwilliam R., Hall P.,
CC RA Grafham D.V., Griffiths C., Griffiths M.N.D., Holden J.L., Howden P.J.,
CC RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Johnson D.,
CC RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
CC RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
CC RA Levaslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
CC RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
CC RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
CC RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
CC RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
CC RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
CC RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
CC RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
CC RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
CC RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
CC RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
CC Rogers J.;
CC RT "The DNA sequence and comparative analysis of human chromosome 20.";
CC RL Nature 414:865-871(2001).
CC RN [3]
CC RP SEQUENCE OF 14-211 FROM N.A. (ISOFORM 1).
CC RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
CC RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
CC RA Nakamura Y., Isono T., Sugano S.;
CC RT "NEDO human cDNA sequencing project.";
CC RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC RN [4]
CC RP SEQUENCE OF 27-211 FROM N.A. (ISOFORM 2).
CC RA Bauer M.;
CC RT "Cloning and sequencing of a new BCAS4 isoform (BCAS4 isoform 1).";
CC RL Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=2;
CC CC Name=1;
CC CC IsoId=Q8TDM0-1; Sequence=Displayed;
CC CC Name=2;
CC CC IsoId=Q8TDM0-2; Sequence=VSP_007854, VSP_007855;
CC CC -!- TISSUE SPECIFICITY: Brain, thymus, spleen, kidney and placenta.
CC CC Overexpressed in most breast cancer cell lines.
CC CC -!- DISEASE: May be involved in breast cancer through a
CC CC t(17;20)(q23;q13) chromosomal translocation that involves BCAS3
CC CC and BCAS4.
CC CC -!- SIMILARITY: Belongs to the cappuccino family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way

```



```
CC -!- FUNCTION: May play a role in hematopoietic lineage decisions and
CC growth regulation. (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- COFACTOR: Magnesium.
CC -!- SUBUNIT: Interacts with MAP3K1.
CC -!- TISSUE SPECIFICITY: Expressed primarily in hematopoietic organs,
CC including bone marrow, spleen and thymus. Also expressed at very
CC low levels in lung, kidney, mammary glands and small
CC intestine.
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC STE20 subfamily.
CC -!- SIMILARITY: Contains 1 CNH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U66464; AAB97983.1; -
CC HSPSP; O63450; IAO6
CC Gnew; HGNC:6863; MAP4K1.
CC MIM; 601983; -
CC GO; GO:0005524; F:ATP binding; IDA.
CC GO; GO:0004674; F:protein serine/threonine kinase activity; IDA.
CC GO; GO:0007257; P:activation of JUNK; TAS.
CC GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
CC GO; GO:0007243; P:protein kinase cascade; IDA.
CC GO; GO:0045610; P:regulation of hemocyte differentiation; ISS.
CC GO; GO:0006950; P:response to stress; IDA.
CC InterPro; IPR001180; Citron.
CC InterPro; IPR000719; Prot kinase.
CC InterPro; IPR008271; Ser Thr pkin AS.
CC InterPro; IPR002290; Ser Thr pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00780; CNH; 1.
CC PRINTS; PR00069; pkinase; 1.
CC PRODOM; PR00109; TYRKINASE.
CC PRODOM; PD000001; Prot kinase; 1.
CC SMART; SM00036; CNH; 1.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN KINASE_ST; FALSE NEG.
CC ATP-binding; Transferase; Kinase; Serine/threonine-protein kinase.
CC DOMAIN 17 274 PROTEIN KINASE.
CC DOMAIN 501 807 CNH.
CC NP_BIND 23 31 ATP (BY SIMILARITY).
CC BINDING 46 46 ATP (BY SIMILARITY).
CC ACT_SITE 137 137 BY SIMILARITY.
CC SEQUENCE 833 AA; 91296 MW; 3C98CF01BE42E151 CRC64;
Query Match 25.0%; Score 62; DB 1; Length 833;
Best Local Similarity 53.3%; Pred. No. 5.3;
Matches 16; Conservative 5; Mismatches 5; Indels 4; Gaps 1;
Qy 23 GSHLHP-----LPGSSDSRASASQASARITGV 48
Db 804 GTISPHCNLLLPGSSNSPASASRVAGITGL 833
RESULT 15
VNCS_PAVEO
ID_VNCS_PAVBO STANDARD; PRT; 726 AA.
AC P07296;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Noncapsid protein NS-1 (Nonstructural protein NS1).
GN NS1.
OS Bovine parvovirus (BPV).
```

```
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10784;
RN SEQUENCE FROM N.A.
RX MEDLINE=87061184; PubMed=3783814;
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,
RA Bates R.C.;
RT "Complete nucleotide sequence and genome organization of bovine
RT parvovirus.";
RL J. Virol. 60:1085-1097(1986).
CC -!- SIMILARITY: Belongs to the parvoviruses noncapsid protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14363; AAB59845.1; -
CC PIR; C26104; UYPVSI.
CC InterPro; IPR001257; Parvo NS1.
CC Pfam; PF01057; Parvo NS1; 1.
CC Nonstructural protein; Noncapsid protein; DNA replication;
CC ATP-binding. 310 317 ATP (POTENTIAL)..
CC NP_BIND 726 AA; 81189 MW; 42659A611A3AF038 CRC64;
CC SEQUENCE 726 AA; 81189 MW; 42659A611A3AF038 CRC64;
Query Match 24.6%; Score 61; DB 1; Length 726;
Best Local Similarity 30.0%; Pred. No. 6;
Matches 12; Conservative 6; Mismatches 22; Indels 0; Gaps 0;
Qy 8 QSFTLSQAGVAWHDGLSLHPLPGSSDSRASASQASARITG 47
Db 486 QNWTFSENGVCVHCGRFQPTPESDTDSGDGDPDPAVAG 525
Search completed: August 10, 2004, 19:14:28
Job time : 14 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: August 10, 2004, 19:11:46 ; Search time 35 seconds
(without alignments)
432.711 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLQSFLLSQAGVAWH.....LPGSSDSRASQSARITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriaph.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	58.1	100	4 Q07826	Q07826 homo sapien
2	135	54.4	109	4 Q8NI81	Q8NI81 homo sapien
3	131	52.8	152	4 Q9NX85	Q9NX85 homo sapien
4	124	50.0	208	4 Q9NW14	Q9NW14 homo sapien
5	122.5	49.4	122	6 Q9BGW3	Q9BGW3 macaca fasc
6	122	49.2	121	4 Q8N8C8	Q8N8C8 homo sapien
7	121	48.8	162	4 Q8NAD5	Q8NAD5 homo sapien
8	118	47.6	158	4 Q8NAL9	Q8NAL9 homo sapien
9	117	47.2	118	4 Q9H387	Q9H387 homo sapien
10	108.5	43.8	138	4 Q8N874	Q8N874 homo sapien
11	106	42.7	157	4 Q8N845	Q8N845 homo sapien
12	105.5	42.5	535	4 Q96EB1	Q96EB1 homo sapien
13	105	42.3	162	4 Q9H5D5	Q9H5D5 homo sapien
14	103	41.5	61	4 Q9UI48	Q9UI48 homo sapien
15	102.5	41.3	135	6 Q95KE1	Q95KE1 macaca fasc
16	102	41.1	133	4 Q96JRS	Q96JRS homo sapien

17	101.5	40.9	375	4	O60448	O60448 homo sapien
18	101	40.7	93	6	O8HXK0	O8HXK0 macaca fasc
19	101	40.7	826	4	O00420	O00420 homo sapien
20	100.5	40.5	180	4	Q8N7M7	Q8N7M7 homo sapien
21	100	40.3	111	4	Q8NSM6	Q8NSM6 homo sapien
22	100	40.3	163	4	Q9GVM0	Q9GVM0 homo sapien
23	100	40.3	175	4	Q8N980	Q8N980 homo sapien
24	98.5	39.7	151	4	Q8N287	Q8N287 homo sapien
25	97.5	39.3	244	4	Q8H700	Q8H700 homo sapien
26	94.5	38.1	171	4	Q9H728	Q9H728 homo sapien
27	94	37.9	180	4	Q8NAQ6	Q8NAQ6 homo sapien
28	92.5	37.3	86	4	Q8NSB5	Q8NSB5 homo sapien
29	92.5	37.3	156	4	Q8NBH4	Q8NBH4 homo sapien
30	92	37.1	170	4	Q8NAI3	Q8NAI3 homo sapien
31	90.5	36.5	1150	4	Q9HBC4	Q9HBC4 homo sapien
32	89.5	36.1	239	4	Q9NX17	Q9NX17 homo sapien
33	89	35.9	140	4	Q8N891	Q8N891 homo sapien
34	89	35.9	156	4	Q8N266	Q8N266 homo sapien
35	88	35.5	179	4	Q8NB20	Q8NB20 homo sapien
36	85.5	34.5	102	4	Q8NAK1	Q8NAK1 homo sapien
37	85.5	34.5	127	4	Q8N8F4	Q8N8F4 homo sapien
38	83.5	33.7	84	4	Q9UHT1	Q9UHT1 homo sapien
39	83	33.5	121	4	Q9HA45	Q9HA45 homo sapien
40	83	33.5	151	4	Q9HA67	Q9HA67 homo sapien
41	82	33.1	126	4	Q9UHT2	Q9UHT2 homo sapien
42	80	32.3	27	4	Q96HX1	Q96HX1 homo sapien
43	79.5	32.1	407	4	O75228	O75228 homo sapien
44	79	31.9	72	4	Q8TB48	Q8TB48 homo sapien
45	79	31.9	79	4	Q9UCY2	Q9UCY2 homo sapien

ALIGNMENTS

RESULT 1

Q07826 PRELIMINARY; PRT; 100 AA.

AC Q07826; (REMBLrel. 01, Created)
DT 01-NOV-1996 (REMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (REMBLrel. 01, Last sequence update)
DE X-linked retinopathy protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93224131; PubMed=8468040;
RA Wong P, MacDonald I M, Sood R, Smith C, Pilon R, Tenniswood M.;
RT "Identification and partial characterization of a candidate gene for
X-linked retinopathies using a lateral approach.";
RL Genomics 15:467-471(1993).
DR EMBL; S58722; AAB26149.1; -.
DR PIR; A46010; A46010.
FT NON_TER 1
SQ SEQUENCE 100 AA; 11065 MW; 50722631CC2CC479 CRC64;

Query Match 58.1%; Score 144; DB 4; Length 100;
Best Local Similarity 62.5%; Pred. No. 2.4e-11;
Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Qy 1 FFFFLQSFLLSQAGVAWHDLGSLHPLPLPGSSDSRASQSARITGV 48

Db 2 FFFFLQSFLLSQAGVAWHDLGSLHPLPLPGSSDSRASQSARITGV 49

RESULT 2

Q8NI81 PRELIMINARY; PRT; 109 AA.

AC Q8NI81; (REMBLrel. 22, Created)
DT 01-OCT-2002 (REMBLrel. 22, Last sequence update)

```

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE OK/SW-CL.41.
GN OK/SW-CL.41.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shchitjo S., Itoh K.;
RT "Identification of immuno-peptidomics that recognized by tumor-reactive
RT CTL generated from TIL of colon cancer patients.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062477; BAB93502.1; -.
SQ SEQUENCE 109 AA; 11580 MW; 15500153CE2B33 CRC64;

Query Match 54.4%; Score 135; DB 4; Length 109;
Best Local Similarity 65.9%; Pred. No. 4.1e-10;
Matches 27; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 8 QSFTLSQAGVAWHDLGSLHPLPGSSDSRASASOSARITGV 48
DB 2 ESHSVTQAGVQVHDLGSLHPLPGSSDSPTASRVAGITGM 42

RESULT 3
Q9NX85
ID Q9NX85 PRELIMINARY; PRT; 152 AA.
AC Q9NX85;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical protein FLJ20378.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Ileal mucosa;
RC Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isgai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000385; BAA91131.1; -.
SQ SEQUENCE 152 AA; 16568 MW; 59065F45AAA301B5 CRC64;

Query Match 52.8%; Score 131; DB 4; Length 152;
Best Local Similarity 38.0%; Pred. No. 2e-09;
Matches 35; Conservative 2; Mismatches 9; Indels 46; Gaps 2;

QY 2 FFFFLRQSTLSQAGVAWHDLGSLHPLPG- 31
DB 23 FFFFLRQSFALVAGVQVHDLGSLQPPGFKQFCLSLSSWDYRHAPCPAYFVFLV 82

QY 32 -----SSDRPASASOSARITG 47
DB 83 DMGPFHVQGTGLELLTSGDPPASASOSAGITG 114

RESULT 4
Q9NW14
ID Q9NW14 PRELIMINARY; PRT; 208 AA.
AC Q9NW14;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein FLJ20837.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adipose tissue;
RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
RA Hira M., Ohmori Y., Ota T., Suzuki Y., Obayashi M., Nishi T.,
RA Shibahara T., Tanaka T., Nakamura Y., Isgai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000844; BAA91396.1; -.
SQ SEQUENCE 208 AA; 23290 MW; C916648B9CAE4520 CRC64;

Query Match 50.0%; Score 124; DB 4; Length 208;
Best Local Similarity 65.7%; Pred. No. 2.5e-08;
Matches 23; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FFFFLRQSTLSQAGVAWHDLGSLHPLPGSSDS 35
DB 137 FFFFLRQSTLSQAGVAWHDLGSLHPLPGSSDS 171

RESULT 5
Q9EGW3
ID Q9EGW3 PRELIMINARY; PRT; 122 AA.
AC Q9EGW3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Frontal cortex;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055293; BAB21918.1; -.
SQ SEQUENCE 122 AA; 13539 MW; 77EFP2E874C50831 CRC64;

Query Match 49.4%; Score 122.5; DB 6; Length 122;
Best Local Similarity 35.2%; Pred. No. 2.1e-08;
Matches 32; Conservative 5; Mismatches 11; Indels 43; Gaps 2;

QY 1 FFFFLR-QSTLSQAGVAWHDLGSLHPLPG- 31
DB 12 FFFFLRQSTLSQAGVAWHDLGSLQPPGFKQFCLSLSSWDYRHAPCPAYFVFLV 82

QY 32 -----SSDRPASASOSARITGV 48
DB 72 SRDYGQAGLELLTSGNPLASASOSAGITGM 102

RESULT 6
Q9N8C8
ID Q9N8C8 PRELIMINARY; PRT; 121 AA.
AC Q9N8C8;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical protein FLJ39655.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```



```
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.,
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK096974; BAC04918.1; -.
KW Hypothetical protein.
SQ SEQUENCE 121 AA; 12917 MW; 4026598DE8735432 CRC64;

Query Match 49.2%; Score 122; DB 4; Length 121;
Best Local Similarity 55.5%; Pred. No. 2.4e-08;
Matches 26; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 3 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDSRASASQARITGV 48
DQ : : : : : : : : : : : : : : : : : : : : : : : :
DB 21 YLFGMEPHVAQAQVQWRDLGSLQPPPPGSRSPASASQVAGITGV 66

RESULT 7
QSNAD5 PRELIMINARY; PRT; 162 AA.
ID QSNAD5
AC QSNAD5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35525.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahori K.,
RA Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092844; BAC03988.1; -.
KW Hypothetical protein.
SQ SEQUENCE 162 AA; 17880 MW; E6D94CE17D0DC71 CRC64;

Query Match 48.8%; Score 121; DB 4; Length 162;
Best Local Similarity 55.3%; Pred. No. 4.6e-08;
Matches 26; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDSRASASQARITG 47
DQ : : : : : : : : : : : : : : : : : : : : : : : :
DB 89 FFFFEHPEHVAQAQVQWRDLGSLQPPPPGSRSPASASQVAGITG 135

RESULT 8
QSNAL9 PRELIMINARY; PRT; 158 AA.
ID QSNAL9
AC QSNAL9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35131.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furiya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yanazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092450; BAC03893.1; -.
KW Hypothetical protein.
SQ SEQUENCE 158 AA; 17414 MW; E151503015F2FE34 CRC64;

Query Match 47.6%; Score 118; DB 4; Length 158;
Best Local Similarity 33.0%; Pred. No. 1.1e-07;
Matches 31; Conservative 6; Mismatches 11; Indels 46; Gaps 2;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDSRASASQARITGV 48
DQ : : : : : : : : : : : : : : : : : : : : : : : :
DB 54 FFFFEKESLSVTQAGVQWRDLGSLQAAAPPPTFFSCLSLPSMNYRRPPCPANFVFL 113

RESULT 9
QSNAD5 PRELIMINARY; PRT; 118 AA.
ID QSNAD5
AC QSNAD5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PRO2550.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
RT "Functional prediction of the coding sequences of 75 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF130089; AAG35515.1; -.
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 13257 MW; 94588870CAC8760D CRC64;

Query Match 47.2%; Score 117; DB 4; Length 118;
Best Local Similarity 36.2%; Pred. No. 1.1e-07;
Matches 34; Conservative 1; Mismatches 13; Indels 45; Gaps 2;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDSRASASQARITGV 48
DQ : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDSRASASQARITGV 98

RESULT 10
QSNAD5 PRELIMINARY; PRT; 138 AA.
ID QSNAD5
AC QSNAD5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ35131.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```

DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ39895.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA Wabe H., Ono T., Hisigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yanashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuna M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097214; BAC04976.1; -.
KW Hypothetical protein.
SQ SEQUENCE 138 AA; 15427 MW; 685645DF92E31994 CRC64;

Query Match 43.8%; Score 108.5; DB 4; Length 138;
Best Local Similarity 34.1%; Pred. No. 1.8e-06;
Matches 31; Conservative 4; Mismatches 11; Indels 45; Gaps 2;

QY 3 FFFLRQSFTL-SQAGVAMHDLGSLHPLPG-----31
DB 13 FLFLWSFTLVAQTGVQWRNLGSLQPPPGPKRPSCLGLPSSWDYRHHAWLIFVLLAEM 72
QY 32 -----SSDSRASASQARITGV 48
DB 73 GFCHVGQAGLELLTSSBPSPASASQAGITGM 103

RESULT 11
Q8N845 PRELIMINARY; PRT; 157 AA.
AC Q8N845;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ40023.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stomach;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yanashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuna M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097342; BAC05007.1; -.
KW Hypothetical protein.
SQ SEQUENCE 157 AA; 17930 MW; D79F77F3B7C608E CRC64;

Query Match 42.7%; Score 106; DB 4; Length 157;
Best Local Similarity 61.3%; Pred. No. 4.4e-06;
Matches 19; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTL-SQAGVAMHDLGSLHPLPG 31
DB 86 FYFFFEESRSVAQAGVQWRDLGSLKPPSPG 116

```

```

RESULT 12
Q85EB1 PRELIMINARY; PRT; 535 AA.
ID Q85EB1;
AC Q85EB1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Similar to hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis, and Embryonic carcinoma;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012514; AAH2514.1; -.
DR Genew; HGNC:1171; ELP4.
DR InterPro; IPR008728; PAXNEB.
DR Pfam; PF05625; PAXNEB; 1.
SQ SEQUENCE 535 AA; 58713 MW; 8656DD3B545B96D4 CRC64;

Query Match 42.5%; Score 105.5; DB 4; Length 535;
Best Local Similarity 57.8%; Pred. No. 2.1e-05;
Matches 26; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

QY 3 FFFLRQSFTL-SQAGVAMHDLGSLHPLPGSSDSRASASQARITG 47
DB 371 FLKLEKFTI-EAGVQWHDLSGRRLGSGGSPASASLVAGITG 414

RESULT 13
Q8H5D5 PRELIMINARY; PRT; 162 AA.
ID Q8H5D5;
AC Q8H5D5;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ23555.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027208; BAB15692.1; -.
KW Hypothetical protein.
SQ SEQUENCE 162 AA; 17632 MW; C9A857907E44D2B8 CRC64;

Query Match 42.3%; Score 105; DB 4; Length 162;
Best Local Similarity 61.3%; Pred. No. 6.1e-06;
Matches 19; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTL-SQAGVAMHDLGSLHPLPG 31
DB 109 FFFFEEMESCSVPAGVQWHDLSGQQPPPG 139

RESULT 14
Q8UI48 PRELIMINARY; PRT; 61 AA.
ID Q8UI48;
AC Q8UI48;
DT 01-MAY-2000 (TReMBLrel. 13, Created)

```

```

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PROC663 (Fragment)
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Yu Y., Zhang C., Luo L., Ouyang S., Zhang S., Li W., Wu J., Zhou S.,
RA Liu M., He F.;
RT "Functional prediction of the coding sequences of 50 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF090944; AAF24056.1; -
FT NON TER 1
SQ SEQUENCE 61 AA; 7364 MW; AAFE987A88277368 CRC64;

Query Match 41.5%; Score 103; DB 4; Length 61;
Best Local Similarity 61.3%; Pred.No.3.7e-06;
Matches 19; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 FFFFLRQSFYLSQAGVAVHDLGSLHPPPG 31
DB 3 FFFFLRQSFYLSQAGVAVHDLGSLHPPPG 33

RESULT 15
Q95KE1 PRELIMINARY; PRT; 135 AA.
AC Q95KE1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Frontal cortex;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062934; BAB60728.1; -
DR GO; GO:0004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . .; IEA.
DR GO; GO:0006096; P:glycolysis; IEA.
DR InterPro; IPR001723; GAP_dhhydrogenase.
DR PROSITE; PS00071; GAPDH; 1.
KW Hypothetical protein.
SQ SEQUENCE 135 AA; 15411 MW; 94DA9FF8025670DF CRC64;

Query Match 41.3%; Score 102.5; DB 6; Length 135;
Best Local Similarity 48.9%; Pred.No.1.1e-05;
Matches 23; Conservative 3; Mismatches 10; Indels 11; Gaps 1;

QY 1 FFFFLRQSFYLSQAGVAVHDLGSLHPPPG 36
DB 77 FFFFLRQSFYLSQAGVAVHDLGSLHPPPG 123

Search completed: August 10, 2004, 19:15:15
Job time : 36 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:10:16 ; Search time 54 Seconds
(without alignments)
251.153 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLRQSTLQAGVAWH.....LPGSSDRASASQARITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248	100.0	48	5	ABG65595 Human bre
2	160	64.5	110	4	AAU33214 Novel hum
3	160	64.5	113	4	AAU32788 Novel hum
4	159.5	64.3	105	4	AAU31652 Novel hum
5	153	61.7	176	7	ADC87115 Human GPC
6	153	61.7	217	7	ADC87151 Human GPC
7	152	61.3	114	4	AAU32787 Novel hum
8	147	59.3	85	4	AAU29593 Novel hum
9	144	58.1	70	3	AB28061 Human sec
10	144	58.1	100	7	ADB37600 Neural th
11	143	57.7	112	4	AAU32786 Novel hum
12	140.5	56.7	194	4	AAU25539 Human pro
13	138.5	55.8	110	4	AAU30350 Novel hum
14	137.5	55.4	52	4	AAU1780 Human pol
15	137.5	55.2	151	7	ADC86929 Human GPC
16	136	54.8	145	4	AAU32025 Novel hum
17	135	54.4	97	5	ABP5149 Human lip
18	135	54.4	103	4	AAU33301 Novel hum
19	135	54.4	103	4	AAU33213 Novel hum
20	135	54.4	109	7	ADD22434 HLA-B46 T
21	135	54.4	164	4	AB117075 Human ner
22	133.5	53.8	153	4	AAU31783 Novel hum
23	133	53.6	142	7	ADB64520 Human pro
24	132	53.2	156	4	ABG10943 Novel hum
25	132	53.2	293	4	AAU71968 Human bon

26	131.5	53.0	71	4	AAU33302	Novel hum
27	131.5	53.0	71	4	AAU33214	Novel hum
28	131.5	53.0	116	4	AAU14736	Novel bon
29	131	52.8	152	7	ADB37629	Neural th
30	131	52.8	164	4	AAU32072	Novel hum
31	131	52.8	229	4	AAU30354	Novel hum
32	130	52.4	53	4	AAO10159	Human pol
33	130	52.4	58	4	AAU95665	Human rep
34	130	52.4	67	4	AAO10596	Human pol
35	130	52.4	70	4	AAO2241	Human pol
36	130	52.4	101	4	AAU32073	Novel hum
37	130	52.4	103	4	AAO8294	Human pol
38	129.5	52.2	104	4	AAU33189	Novel hum
39	129.5	52.2	104	4	AAU33156	Novel hum
40	129.5	52.2	104	4	AAU33281	Novel hum
41	129	52.0	47	4	AAO10724	Human pol
42	129	52.0	117	7	ADD19326	Human sec
43	129	52.0	117	7	ADD19292	Human sec
44	129	52.0	138	4	AAU32881	Novel hum
45	128.5	51.8	324	4	AAU29573	Novel hum

ALIGNMENTS

RESULT 1
ABG65595
ID ABG65595 standard; protein; 48 AA.

AC ABG65595;
XX
DT 23-AUG-2002 (first entry)
XX
DE Human breast specific polypeptide, BSP #65.

XX
KW Human; breast specific protein; BSP; breast cancer; vaccine;
KW mammary tumour; cytostatic; gene therapy; non-cancerous breast disorder.

XX Homo sapiens.

XX WO200236807-A2.

PD 10-MAY-2002.

PF 29-OCT-2001; 2001WO-US046888.

PR 27-OCT-2000; 2000US-0243802P.

XX (DIAD-) DIADEXUS INC.

PI Sun Y, Recipon H, Salceda S, Liu C, Turner LR;

DR WPI; 2002-463415/49.

PT New breast-specific nucleic acids and polypeptides, useful for
PT identifying, diagnosing, monitoring, staging, imaging, and treating
PT breast cancer and non-cancerous disease states in breast tissues.

FS Claim 11; Page 245; 281pp; English.

XX The invention relates to breast specific nucleic acids (BSNA) and breast-specific polypeptides (BSP). Also included are a method for determining the BSNA in a sample, a vector comprising a BSNA, a host cell comprising the BSNA in a sample, a method for producing a polypeptide encoded by a BSNA, an anti-BSP antibody and a method for determining the presence of a BSP in a sample. The breast-specific nucleic acids, polypeptides and compositions comprising them are useful for identifying, diagnosing, monitoring, staging, imaging, and treating breast cancer, mammary tumor and non-cancerous disease states in breast tissue; for identifying breast tissue; for monitoring, identifying and/or designing agonists and antagonists of the polypeptides; in gene therapy; in producing transgenic animals and cells; for producing engineered breast tissue for treatment and research; and as elements in an array or computer program for pattern recognition

CC of breast disorders. The nucleic acids may be used as hybridisation
 CC probes to detect, characterise and quantify hybridising nucleic acids in,
 CC and isolate hybridising nucleic acids from, both genomic and transcript-
 CC derived nucleic acid samples. The BSP protein may be used in a vaccine
 CC composition for raising an immune response against breast cancer. The
 CC present sequence is BSP protein of the invention
 XX
 SQ Sequence 48 AA;

Query Match 100.0%; Score 248; DB 5; Length 48;
 Best Local Similarity 100.0%; Pred. No. 3.4e-25;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFFFLRQFTLSQAGVAMHDLGSLHPPPLPGSSDSRASASQARITGV 48
 DB 1 FFFFLRQFTLSQAGVAMHDLGSLHPPPLPGSSDSRASASQARITGV 48

RESULT 2

AAU31552
 ID AAU31552 standard; protein; 110 AA.

XX AC AAU31552;

DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #2043.

XX KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US008656.

XX PR 18-APR-2000; 2000US-00552929.

XX PR 26-JAN-2001; 2001US-00770160.

XX XX (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-611725/70.

XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 XX vaccination, testing and therapy.

PS Claim 20; Page 457; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention

XX SQ Sequence 110 AA;

Query Match 64.5%; Score 160; DB 4; Length 110;
 Best Local Similarity 69.6%; Pred. No. 3.6e-13;
 Matches 32; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 2 FFFFLRQFTLSQAGVAMHDLGSLHPPPLPGSSDSRASASQARITGV 47
 DB 1 FFFFEVESRTVTQAGVQWHDLGSLQPPPPGSSDSPVSASHVAEITG 46

RESULT 3

AAU32788

ID AAU32788 standard; protein; 113 AA.

XX AC AAU32788;

DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #3279.

XX KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US008656.

XX PR 18-APR-2000; 2000US-00552929.

XX PR 26-JAN-2001; 2001US-00770160.

XX XX (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-611725/70.

XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 XX vaccination, testing and therapy.

PS Claim 20; Page 663-664; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention

XX SQ Sequence 113 AA;

Query Match 64.5%; Score 160; DB 4; Length 113;
 Best Local Similarity 69.6%; Pred. No. 3.7e-13;
 Matches 32; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 FFFFLRQFTLSQAGVAMHDLGSLHPPPLPGSSDSRASASQARITGV 47
 DB 31 FFFETESCSVAQAGVQWHDLGSLQPPPPGSSDSPASASVAGITG 76

```
RESULT 4
AAU31652
ID AAU31652 standard; protein; 105 AA.
XX
AC AAU31652;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secreted protein #2143.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
FN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
EF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
XX
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Dmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 492; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention.
XX
SQ Sequence 105 AA;
Query Match 64.3%; Score 159.5; DB 4; Length 105;
Best Local Similarity 40.9%; Pred. No. 4e-13;
Matches 38; Conservative 2; Mismatches 8; Indels 45; Gaps 1;
QY 1 FFFFLRQSTLSQAGVAWHDLSLHPPLPG----- 31
DB 1 FFFFLRQSFALQAQGWHDLSLQPPPGFKXFSYLSPLSSWDYRYVPPRPASPEFLV 60
QY 32 -----SSDSRASASQARITGV 48
DB 61 EMGFHVHGQAGFELLTSSDPPASASQAGIIGV 93
RESULT 5
ADC87115
ID ADC87115 standard; protein; 176 AA.
XX
AC ADC87115;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1568.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
FN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PR 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR N-PSDB; ADC87114.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 2; SEQ ID NO 1568; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
XX
SQ Sequence 176 AA;
Query Match 61.7%; Score 153; DB 7; Length 176;
Best Local Similarity 66.0%; Pred. No. 5.1e-12;
Matches 31; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 FFFFLRQSTLSQAGVAWHDLSLHPPLPGSSDSRASASQARITG 47
DB 4 FFFFLRQSFHVTQAGVQWCMNFSSLPPLPGSSDSPASASQVAGITG 50
RESULT 6
ADC87151
ID ADC87151 standard; protein; 217 AA.
XX
AC ADC87151;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR protein SEQ ID NO:1604.
XX
KW human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
FN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
```

PF 18-JUN-2002.; 2002EP-00013517.
 XX
 PR 18-JUN-2001.; 2001JP-00246789.
 XX
 XX (NRAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 PI WPI; 2003-315783/31.
 XX N-PSDB; ADC87150.
 DR
 DR
 XX
 PT New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX
 XX Claim 2; SEQ ID NO 1604; 28pp; English.
 PS
 CC The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
 XX
 XX Sequence 217 AA;
 SQ
 Query Match 61.7%; Score 153; DB 7; Length 217;
 Best Local Similarity 68.1%; Pred. No. 6.5e-12;
 Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 PFFFLRQSFTLSQAGVAWHDLGSLHPPLPGSSDSRASASQSRITG 47
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 146 PFFFFETESHTVTWAGVQWCDLGLSLQPPPPGSSDSPASASRVAGITG 192
 RESULT 7
 ID AAU32787 standard; protein; 114 AA.
 XX
 AC AAU32787;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #3278.
 XX
 XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 PD
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 PT
 PS Claim 20; Page 663; 765pp; English.
 XX

CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation, to regulate haematopoiesis, and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU3304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 XX Sequence 114 AA;
 SQ
 Query Match 61.3%; Score 152; DB 4; Length 114;
 Best Local Similarity 66.7%; Pred. No. 4.3e-12;
 Matches 32; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
 QY 1 PFFFLRQSFTLSQAGVAWHDLGSLHPPLPGSSDSRASASQSRITGV 48
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 Db 38 PFFFFETESHCHVAGVQWDLGLSLQPPPPGSSNSPASASQVACTTGV 85
 RESULT 8
 ID AAU29593 standard; protein; 85 AA.
 XX
 AC AAU29593;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #84.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 PD
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 PR 26-JAN-2001; 2001US-00770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 PT
 PS Claim 20; Page 168; 765pp; English.
 XX
 XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the

CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX Sequence 85 AA;

Query Match 59.3%; Score 147; DB 4; Length 85;
 Best Local Similarity 62.5%; Pred. No. 1.4e-11;
 Matches 30; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDRASASQSARITGV 48
 Db 5 YLFIVTESCTVVGQVQWCDLGLQLPPLPGSSDPHASTSXVAGITGV 52

RESULT 9

AA28061
 ID AAB28061 standard; protein; 70 AA.

XX AAB28061;

DT 02-FEB-2001 (first entry)

DE Human secreted protein BLAST search protein SEQ ID NO: 109.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW candiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO2000055177-A2.

XX 21-SEP-2000.

XX 09-MAR-2000; 2000WO-US006058.

XX 12-MAR-1999; 99US-0124145P.

XX 03-DEC-1999; 99US-0168654P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-638177/61.

XX Novel nucleic acids encoding 49 human secreted proteins useful for
 PT treating cancers, hyperproliferative disorders, inflammatory disorders,
 PT neurological disorders and cardiovascular disorders.

XX Disclosure; Page 364-365; 389pp; English.

XX The invention relates to the isolation of genes AAC59108-C59156 encoding
 CC the human secreted proteins AAB28012-B28060. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene given in
 CC the descriptor line. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections
 XX Sequence 70 AA;

Query Match 58.1%; Score 144; DB 3; Length 70;
 Best Local Similarity 62.5%; Pred. No. 2.8e-11;
 Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAWHDLGSLHPLPGSSDRASASQSARITGV 48
 Db 1 FFFFPETESCSVAEAGVQWCDLGLSKSPPPGSSDSPASASRVAGITGM 48

RESULT 10

ADB37600

ID ADB37600 standard; protein; 100 AA.

XX ADB37600;

DT 04-DEC-2003 (first entry)

XX Neural thread protein-related protein #63.

XX Cytostatic; Antitumour; Antipsoriatic; Dermatological;

KW Antiatherosclerotic; Antiarteriosclerotic; Vasotropic; Antiinflammatory;

KW Immunosuppressive; Tranquillizer; Antiemetic; Virucide; AD7C-NTP;

KW neural thread protein; neuritic sprouting.

XX Unidentified.

XX WO2003008444-A2.

XX 30-JAN-2003.

XX 19-JUL-2002; 2002WO-CA001106.

XX 19-JUL-2001; 2001US-0306150P.

XX 19-JUL-2001; 2001US-0306161P.

XX 16-NOV-2001; 2001US-0331477P.

XX (NYMO-) NYMOX CORP.

XX Averbach PA, Gemmell J;

XX WPI; 2003-248000/24.

XX Novel Related peptide or AD7c-neural thread peptide, useful for treating
 PT unwanted cellular proliferations, glandular hyperplasia, unwanted facial
 PT hair, warts and unwanted fatty tissue.

XX Claim 1; Page 45; 109pp; English.

XX The present invention relates to AD7c-neural thread protein (NTP) and
 CC related proteins and peptides (I; ADB37528-ADB37641). The sequences are
 CC useful for treating a condition in a patient requiring removal or
 CC destruction of cells. The condition can be selected from benign or
 CC malignant tumour of a tissue, hyperplasia, hypertrophy or overgrowth of a
 CC tissue, vitally, bacterially or parasitically altered tissue, or
 CC malformation of a tissue, where the tissue is selected from lung, breast,
 CC stomach, pancreas, prostate, bladder, bone, ovary, skin, kidney, sinus,
 CC colon, intestine, stomach, rectum, oesophagus, heart, spleen, salivary
 CC gland, blood, brain and its coverings, spinal cord and its coverings,
 CC muscle, connective tissue, adrenal, parathyroid, thyroid, uterus, testis,
 CC pituitary, reproductive organs, liver, gall bladder, eye, ear, nose,
 CC throat, tonsils, mouth, lymph nodes and lymphoid tissue. The condition is
 CC preferably tonsillary hypertrophy, prostatic hyperplasia, psoriasis,
 CC eczema, dermatosis, cosmetic modification to a tissue (skin, eye, ear,
 CC nose, throat, mouth, muscle, connective, hair or breast tissue), vascular
 CC disease (atherosclerosis or arteriosclerosis), haemorrhoids, varicose

CC veins, inflammatory disease, autoimmune disease, metabolic disease,
 CC hereditary/genetic disease, traumatic disease or physical injury,
 CC nutritional deficiency disease, infectious disease, amyloid disease,
 CC fibrosis disease, storage disease, congenital malformation, enzyme
 CC deficiency disease, poisoning, intoxication, environmental disease,
 CC radiation disease, endocrine disease, degenerative disease and mechanical
 CC disease. The peptides are useful for treating unwanted cellular
 CC proliferations, glandular (e.g. prostate) hyperplasia, unwanted facial
 CC hair, warts and unwanted fatty tissue, or for preparing antibodies that
 CC recognize and/or bind to related proteins, Related peptides or NTP
 CC peptides.
 XX
 SQ Sequence 100 AA;

Query Match 58.1%; Score 144; DB 7; Length 100;
 Best Local Similarity 62.5%; Pred. No. 4.2e-11;
 Matches 30; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
 QY 1 FFFFLRQFTLSQAGVAVHDLGSLHPLPGSSDSRASASQSRITGV 48
 DB 2 FFFFEETESCSVAEAGVQWCDLGLSKSPPGSSDSRASASRVAGITGM 49

RESULT 11
 AAU32786
 ID AAU32786 standard; protein; 112 AA.
 XX
 AC AAU32786;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #3277.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US008656.
 XX
 PR 18-APR-2000; 2000US-00552929.
 XX
 PR 26-JAN-2001; 2001US-00770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.
 XX
 PS Claim 20; Page 663; 765pp; English.
 XX

CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the
 CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or

CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention
 XX
 SQ Sequence 112 AA;

Query Match 57.7%; Score 143; DB 4; Length 112;
 Best Local Similarity 65.9%; Pred. No. 6.4e-11;
 Matches 29; Conservative 6; Mismatches 9; Indels 0; Gaps 0;
 QY 4 FFLRQFTLSQAGVAVHDLGSLHPLPGSSDSRASASQSRITG 47
 DB 1 FFEETESNSVAQAGVQWHDGLSLQPPSPGSSNSPASASRVAGITG 44

RESULT 12
 AAU25539
 ID AAU25539 standard; protein; 194 AA.
 XX
 AC AAU25539;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:1054.
 XX
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiac; antidiabetic; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnary; antileukemic; cytostatic;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antitense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX Homo sapiens.
 OS
 PN WO200153455-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 22-DEC-2000; 2000WO-US035017.
 XX
 PR 23-DEC-1999; 99US-00471275.
 XX
 PR 21-JAN-2000; 2000US-00488725.
 XX
 PR 25-APR-2000; 2000US-00552317.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-457603/49.
 XX
 DR N-PSDB; AAU99480.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
 XX
 PS Claim 20; Page 212; 1217pp; English.

CC AAU99166 to AAU99904 encode the human proteins given in AAU25225 to
 CC AAU25963. The proteins can have activities based on the tissues and cells
 CC they are expressed in, such as: antineoplastic; antirheumatic;
 CC antiallergic; immunosuppressive; antidiabetic; endocrine; cardiac;
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
 CC antileukemic; osteopathic; dermatological; antiallergic; antiasthmatic;
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;

CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders
 XX
 SQ Sequence 194 AA;

Query Match 56.7%; Score 140.5; DB 4; Length 194;
 Best Local Similarity 39.6%; Pred. No. 2.6e-10;
 Matches 36; Conservative 1; Mismatches 9; Indels 45; Gaps 2;

QY 3 FFFLRQFTLL-SQAGVAVHDLGSLHPPPLPG----- 31
 DB 9 FFFLRWSFALVAQAGVQWHDLSGLQPPAPGKRFSSLSLSLRWDYRHAHARLIFVFLVEM 68
 QY 32 -----SSDSRASASQSARITGV 48
 DB 69 GFLHVGAGLELPTSGDPPTTSASQSARITGV 99

RESULT 13

AAU30350
 ID AAU30350 standard; protein; 110 AA.

XX
 AC AAU30350;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #841.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.

XX
 EN WO200179449-A2.

XX
 PD 25-OCT-2001.

XX
 PF 16-APR-2001; 2001WO-US008656.

XX
 PR 18-APR-2000; 2000US-00552929.

FR 26-JAN-2001; 2001US-00770160.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT;

XX
 DR WPI; 2001-611725/70.

XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy.

XX
 PS Claim 20; Page 281; 765pp; English.

XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising the

CC nucleic acids encoding the polypeptides and cells genetically engineered
 CC to express them are also useful for producing the proteins. The proteins
 CC are useful in genetic vaccination, testing and therapy, and can be used
 CC as nutritional supplements. They may be used to increase stem cell
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
 CC and/or nerve tissue growth or regeneration; immune suppression and/or
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human
 CC secreted proteins of the invention

XX
 SQ Sequence 110 AA;

Query Match 55.8%; Score 138.5; DB 4; Length 110;
 Best Local Similarity 58.5%; Pred. No. 2.5e-10;
 Matches 31; Conservative 5; Mismatches 12; Indels 5; Gaps 1;

QY 1 FFFFL-----RQFTLSQAGVAVHDLGSLHPPPLPGSSDSRASASQSARITGV 48
 DB 13 FFFFLFFFTETESCPVAQAGTQWCDLSGLQPPPGXNSPASASRVAGITGV 65

RESULT 14

AAO11780

ID AAO11780 standard; protein; 52 AA.

XX
 AC AAO11780;

XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human polypeptide SEQ ID NO 25672.

XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX
 FN WO200164835-A2.

XX
 PD 07-SEP-2001.

XX
 PF 26-FEB-2001; 2001WO-US004927.

XX
 PR 28-FEB-2000; 2000US-00515126.

FR 18-MAY-2000; 2000US-00577405.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT;

XX
 DR WPI; 2001-514838/56.

DR N-PSDB; AAI91711.

XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.

XX
 PS Claim 20; SEQ ID NO 25672; 1399pp + Sequence Listing; English.

XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at fcp.wipo.int/pub/published_pct_sequences

XX

Search completed: August 10, 2004, 19:14:04
Job time : 55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:15:22 ; Search time 46 Seconds
(without alignments)
327.321 Million cell updates/sec

Title: US-10-082-830-224

Perfect score: 248

Sequence: 1 FFFFLRQSFLLSQAGVAMH.....LFGSSDRASASQARITGV 48

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	248	100.0	48	14	US-10-082-830-224
2	153	61.7	176	14	Sequence 224, App
3	153	61.7	176	15	Sequence 1912, Ap
4	153	61.7	217	14	Sequence 1568, Ap
5	153	61.7	217	14	Sequence 1956, Ap
6	144	58.1	217	15	Sequence 1604, Ap
7	140.5	56.7	194	12	Sequence 82, Appl
8	137	55.2	129	15	Sequence 1054, Ap
9	137	55.2	151	14	Sequence 4804, Ap
10	137	55.2	151	15	Sequence 1726, Ap
11	133	53.6	142	15	Sequence 1382, Ap
12	131	52.8	152	14	Sequence 2674, Ap
13	130	52.4	58	10	Sequence 112, App
14	126	50.8	220	14	Sequence 4323, Ap
15	126	50.8	220	15	Sequence 1698, Ap
					Sequence 1354, Ap

16	125	50.4	53	12	US-10-276-774-1962	Sequence 1962, Ap
17	124	50.0	208	14	US-10-198-070-115	Sequence 115, App
18	124	50.0	208	14	US-10-176-306-56	Sequence 56, Appl
19	124	50.0	256	12	US-10-236-417-68	Sequence 68, Appl
20	122	49.2	121	15	US-10-108-260A-4053	Sequence 4053, Ap
21	121	48.8	162	15	US-10-104-047-3256	Sequence 3256, Ap
22	119	48.0	94	13	US-10-011-445-58	Sequence 58, Appl
23	119	48.0	102	12	US-10-221-278-674	Sequence 674, App
24	119	48.0	102	15	US-10-291-172-674	Sequence 3047, App
25	118	47.6	158	15	US-10-104-047-3047	Sequence 30, Appl
26	117	47.2	118	14	US-10-198-070-30	Sequence 1734, Ap
27	117	47.2	190	14	US-10-017-161-1734	Sequence 1390, Ap
28	117	47.2	190	15	US-10-232-798-1390	Sequence 1865, App
29	116.5	47.0	133	15	US-10-094-749-1865	Sequence 908, App
30	116	46.8	97	12	US-10-112-944-908	Sequence 1859, Ap
31	115	46.4	137	12	US-10-276-774-1959	Sequence 4725, Ap
32	114	46.0	109	15	US-10-108-260A-4725	Sequence 1950, Ap
33	113	45.6	146	12	US-10-276-774-1950	Sequence 18, Appl
34	112.5	45.4	124	15	US-10-144-1948-18	Sequence 5341, Ap
35	112	45.2	68	10	US-09-764-891-5341	Sequence 107, App
36	111.5	45.0	361	9	US-09-995-494-107	Sequence 1940, Ap
37	111	44.8	157	12	US-10-276-774-1940	Sequence 1980, Ap
38	110.5	44.6	118	12	US-10-276-774-1980	Sequence 2008, Ap
39	110	44.4	97	12	US-10-276-774-2008	Sequence 79, Appl
40	110	44.4	115	9	US-09-995-494-79	Sequence 3250, Ap
41	110	44.4	176	15	US-10-108-260A-3250	Sequence 3685, Ap
42	109.5	44.2	149	15	US-10-104-047-3685	Sequence 2884, Ap
43	109.5	44.2	273	15	US-10-108-260A-2884	Sequence 2092, Ap
44	109.5	44.2	433	14	US-10-017-161-2092	Sequence 1738, Ap
45	109.5	44.2	433	15	US-10-292-798-1738	

ALIGNMENTS

RESULT 1

US-10-082-830-224
; Sequence 224, Application US/10082830
; Publication No. US20030077604A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Salceda, Susana
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0249
; CURRENT APPLICATION NUMBER: US/10/082,830
; CURRENT FILING DATE: 2001-10-29
; PRIOR FILING DATE: 60/243,802
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 224
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-082-830-224

Query Match 100.0%; Score 248; DB 14; Length 48;
Best Local Similarity 100.0%; Pred. No. 6.1e-24;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFFFLRQSFLLSQAGVAMHDLGSLHPLPGSSDRASASQARITGV 48
DB 1 FFFFLRQSFLLSQAGVAMHDLGSLHPLPGSSDRASASQARITGV 48

RESULT 2

US-10-017-161-1912
; Sequence 1912, Application US/10017161
; Publication No. US20030143668A1

GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1912
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1912

Query Match 61.7%; Score 153; DB 14; Length 176;
Best Local Similarity 66.0%; Pred. No. 2.7e-11;
Matches 31; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPPLPGSSDSRASQSARITG 47
||||| : : : : : : : : : : : : : : : : : :
DB 4 FFFFEESHFTVQAGVQWCMNFSSLPPLPGSSDSPASASQVAGITG 50

RESULT 3

US-10-292-798-1568
; Sequence 1568, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1568
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1568

Query Match 61.7%; Score 153; DB 15; Length 176;
Best Local Similarity 66.0%; Pred. No. 2.7e-11;
Matches 31; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPPLPGSSDSRASQSARITG 47
||||| : : : : : : : : : : : : : : : : : :
DB 4 FFFFEESHFTVQAGVQWCMNFSSLPPLPGSSDSPASASQVAGITG 50

RESULT 4

US-10-017-161-1956
; Sequence 1956, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152

; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1956
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1956

Query Match 61.7%; Score 153; DB 14; Length 217;
Best Local Similarity 68.1%; Pred. No. 3.4e-11;
Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPPLPGSSDSRASQSARITG 47
||||| : : : : : : : : : : : : : : : : : :
DB 146 FFFFEESHFTVQAGVQWCDLGLSLQPPPGSSDSPASASRVAGITG 192

RESULT 5

US-10-292-798-1604
; Sequence 1604, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1604
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1604

Query Match 61.7%; Score 153; DB 15; Length 217;
Best Local Similarity 68.1%; Pred. No. 3.4e-11;
Matches 32; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 FFFFLRQSFTLSQAGVAMHDLGSLHPPPLPGSSDSRASQSARITG 47
||||| : : : : : : : : : : : : : : : : : :
DB 146 FFFFEESHFTVQAGVQWCDLGLSLQPPPGSSDSPASASRVAGITG 192

RESULT 6

US-10-198-070-82
; Sequence 82, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; APPLICANT: GENWELL, JACK
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 59003.000008
; CURRENT APPLICATION NUMBER: US/10/198,070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477

```

; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
; LENGTH: 100
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-198-070-82

```

Query Match 58.1%; Score 144; DB 14; Length 100;
Best Local Similarity 62.5%; Pred. No. 2e-10;
Matches 30; Conservative 7; Mismatches 11; Indels

QY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
QY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
db	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

RESULT 7

```

US-10-296-115-1054
; Sequence 1054, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hvseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1054
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-296-115-1054

```

Query Match 56.7%; Score 140.5; DB 12; Length 194;
Best Local Similarity 39.6%; Pred. No. 1.1e-09;
Matches 36: Conservative 1; Mismatches 9; Indels 45; Gaps 2;

```
QY      3 FFFLRQSFTL-SQAGVAVHDLCSLHPPLPG----- 31
          ||||| : ||||| ||||| |||||
Rb      9 PPEI DISEPATVAACGCAWMDHLAST-ODDAPCGKPKSSSTLSPPDYRHHARLI.FEVLIVEM 68
```

Qy	32	-----SSDSRASASQARITGV	48
Db	59	CEIHLVGCAGCTETDPTSCDDPPTCASASQARITGV	99

RESULTS

```

RESUL 8
US-10-108-260A-4804
; Sequence 4804, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4804
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4804

```

Query Match	55.2%;	Score 137;	DB 15;	Length 129;
Best Local Similarity	62.5%;	Pred. No. 2e-09;		
Matches 30: Conservative	1: Mismatches	17: Indels		

QY 1 FFFFLRQSFTLSQAGVAWHDLSLHPLPGSSDSRASASQSARITGV 48
||||| : |||||
pb 53 FFFFLDGVLLCRGVGVWRNLGSLPAPCGSSDASAASQVARTTDV 100

RESULT 9

US-10-017-161-1726
; Sequence 1726, Application US/10017161
; Publication No. US2003014368A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ASURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1726
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-017-161-1726

Query Match	55.2%	Score 137;	DB 14;	Length 151;
Best Local Similarity	61.1%;	Pred. No. 2.4e-09;		
Matches 33;	Conservative	4;	Mismatches 11;	Indels 6;
				Gaps 2;

QY
1 FFFFLRQSFTLSQAGVAKHDLGSLH--PP----LPGSDSRASASQARITGV 48
||| ||| : ||| ||| ||| ||| ||| ||| : ||| :
nD
18 FFVFPPVOSHVSVOAGVOWHDLGSKOSLPSPNRLPGSSDSPASARVAGITGM 71

RESIST 10

RESULT 10
 US-10-292-798-1382
 ; Sequence 1382, Application US/10292798
 ; Publication No. US20030235833a1
 ; GENERAL INFORMATION:
 ; APPLICANT: SUWA, MAKIKO
 ; APPLICANT: ASAI, KIYOSHI
 ; APPLICANT: AKIYAMA, YUTAKA
 ; APPLICANT: ABURATANI, HIROYUKI
 ; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
 ; FILE REFERENCE: 084335/166
 ; CURRENT APPLICATION NUMBER: US/10/292,798
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: 10/017,151
 ; PRIOR FILING DATE: 2001-12-18
 ; PRIOR APPLICATION NUMBER: JP 2001-246789
 ; PRIOR FILING DATE: 2001-06-18
 ; NUMBER OF SEQ ID NOS: 2070
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1382
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-292-798-1382

Query Match	55.2%	Score 137;	DB 15;	Length 151;
Best Local Similarity	61.1%	Pred. NO. 2.4e-09;		
Matches 13.	Conservative	4:	Mismatches 11;	Indels 6;
				Gaps 2

Qy 1 FFFFLRQSFTLSQAGVAWHDLSLH--PP---LPSSDSRASASQSRITGV 48
||| ||| :: ||| ||| ||| ||| ||| ||| ||| :
Db 18 FFVFFVQSHSVVOAGVOWHDIGSKOSLPPSNLPLPGSSDGPASASRVAGITGM 71

RESULT 11

```

US-10-104-047-2674
; Sequence 2674, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104.047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2674
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2674

Query Match      53.6%; Score 133; DB 15; Length 142;
Best Local Similarity 54.2%; Pred. No. 7.2e-09;
Matches 26; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

Qy 1 FFFFLRQSTLSQAGVAMHDLGSLHPPPLGSSDSRASASQARITGV 48
Db 43 FLFLETESHVAGAGVQWHDLSLQPTPTSSNSPASVFRVAEIIQM 90

RESULT 12
US-10-198-070-112
; Sequence 112, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; FILE REFERENCE: 59003.000008
; CURRENT APPLICATION NUMBER: US/10/198.070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306.161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306.150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331.477
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 112
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-198-070-112

Query Match      52.8%; Score 131; DB 14; Length 152;
Best Local Similarity 38.0%; Pred. No. 1.4e-08;
Matches 35; Conservative 2; Mismatches 9; Indels 46; Gaps 2;

Qy 2 FFFFLRQSTLSQAGVAMHDLGSLHPPPLG----- 31
Db 23 FFFFLRQSFALVAQAGVQWRNLSLQPPPPGKQFSCLLSLSDWYRIAPPCAYFVFLV 82
Qy 32 -----SSDSRASASQARITG 47
Db 83 DMGFPHVGTGLLELTSGDPPASASQAGITG 114

RESULT 13
US-09-764-891-4323
; Sequence 4323, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: ROSSEN ET AL.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND ANTIBODIES
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764.891
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - CONSULT PALM OR FILE WRAPPER
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4323
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (45)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (53)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4323

Query Match      52.4%; Score 130; DB 10; Length 58;
Best Local Similarity 71.4%; Pred. No. 6.4e-09;
Matches 25; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 11 TLSQAGVAMHDLGSLHPPPLGSSDSRASASQASARI 45
Db 4 SVTQAGVQWNLGSLQPLPGSSDSPASASQVAGV 38

RESULT 14
US-10-017-161-1698
; Sequence 1698, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017.161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1698
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-161-1698

Query Match      50.8%; Score 126; DB 14; Length 220;
Best Local Similarity 56.5%; Pred. No. 9e-08;
Matches 26; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

Qy 2 FFFFLRQSTLSQAGVAMHDLGSLHPPPLGSSDSRASASQARITG 47
Db 44 YFFFLRQGLTVAQAGVQWCMHNSLQPKTGLSLNPPTSASQVAMTTG 89

RESULT 15
US-10-292-798-1354
; Sequence 1354, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA

```


Sun Aug 15 07:51:01 2004

```
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1354
; LENGTH: 220
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-798-1354

Query Match      50.8%; Score 126; DB 15; Length 220;
Best Local Similarity 56.5%; Pred. No. 9e-08;
Matches 26; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY      2 FFFFLRQSFTLSQAGVAWHDLSLHPPLPGSSDSRASASQSRITG 47
      :||||| :||| :||| :||| :||| :||| :||| :|||
Db      44 YFFFLRQGLTVAQAGVQWCHGSLQPKTPGLSNPPTASQVAMTTG 89
```

Search completed: August 10, 2004, 19:20:51
Job time : 47 secs

This Page Blank (uspto)